VOLUME 103 PART C NUMBER 3 MARCH 1956

The Proceedings OF THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART C

MONOGRAPHS Nos. 132–158

SAVOY PLACE · LONDON W.C.2

Price Eight Shillings and Sixpence

The Institution of Electrical Engineers

FOUNDED 1871
INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1955–1956

President

SIR GEORGE H. NELSON, BART.

Past-Presidents

SIR JAMES SWINBURNE, BART., F.R.S.
W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGCUMBE, T.D.
J. M. DONALDSON, M.C.
PROFESSOR E. W. MARCHANT, D.Sc.
P. V. HUNTER, C.B.E.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
SIR ARTHUR P. M. FLEMING, C.B.E., D.Eng., LL.D.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).
COLONEL SIR A. STANLEY ANGWIN, K.B.E., D.S.O., M.C., T.D., D.Sc.(Eng.).
SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc.(Eng.).
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROFESSOR E. B. MOULLIN, M.A., Sc.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COLONEL B. H. LEESON, C.B.E., T.D.
SIR HAROLD BISHOP, C.B.E., B.Sc.(Eng.).
J. ECCLES, C.B.E., B.Sc.

Vice-Presidents

T. E. GOLDUP, C.B.E.
S. E. GOODALL, M.Sc.(Eng.).
WILLIS JACKSON, D.Sc., D.Phil., Dr.Sc.Tech., F.R.S.
SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C., LL.D., B.Sc.
SIR W. GORDON RADLEY, C.B.E., Ph.D.(Eng.).

Honorary Treasurer

THE RT. HON. THE VISCOUNT FALMOUTH.

Ordinary Members of Council

PROFESSOR H. E. M. BARLOW, Ph.D., B.Sc.(Eng.).
J. BENNETT.
C. M. COCK.
A. R. COOPER, M.Eng.
A. T. CRAWFORD, B.Sc.
B. DONKIN, B.A.
PROFESSOR J. GREIG, M.Sc., Ph.D.
F. J. LANE, O.B.E., M.Sc.
G. S. C. LUCAS, O.B.E.
D. McDONALD, B.Sc.
C. T. MELLING, C.B.E., M.Sc.Tech.
H. H. MULLENS, B.Sc.
W. F. PARKER.
R. L. SMITH-ROSE, C.B.E., D.Sc., Ph.D.
G. L. WATES, J.P.
G. O. WATSON.
D. B. WELBOURN, M.A.
J. H. WESTCOTT, B.Sc.(Eng.), Ph.D.
E. L. E. WHEATCROFT, M.A.
R. T. B. WYNN, C.B.E., M.A.

Chairman and Past-Chairmen of Sections

Measurement and Control:
W. BAMFORD, B.Sc.
*M. WHITEHEAD.

Radio and Telecommunication:
H. STANESBY.
*C. W. OATLEY, O.B.E., M.A., M.Sc.

Supply:
L. DRUCQUER.
*J. D. PEATTIE, B.Sc.

Utilization:
D. B. HOGG, M.B.E., T.D.
*J. I. BERNARD, B.Sc.Tech.

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:
F. R. C. ROBERTS.
*J. M. MITCHELL, B.Sc., Ph.D.

Mersey and North Wales Centre:
PROFESSOR J. M. MEEK, D.Eng.
*P. R. DUNN, B.Sc.

North Midland Centre:
F. BARRELL.
*G. CATON.

North-Eastern Centre:
A. H. KENYON.
*G. W. B. MITCHELL, B.A.

North-Western Centre:
G. V. SADLER.
*PROFESSOR E. BRADSHAW, M.B.E., M.Sc.Tech., Ph.D.

Northern Ireland Centre:
MAJOR E. N. CUNLIFFE, B.Sc.Tech.
*MAJOR P. L. BARKER, B.Sc.

Western Centre:
T. G. DASH, J.P.
*A. N. IRENS.

Scottish Centre:
*E. WILKINSON, Ph.D., B.Eng.
*J. S. HASTIE, B.Sc.(Eng.).

South Midland Centre:
H. S. DAVIDSON, T.D.
*A. R. BLANDFORD.

Southern Centre:
L. H. FULLER, B.Sc.(Eng.).
*E. A. LOGAN, M.Sc.

* Past-Chairman.

Secretary

W. K. BRASHER, C.B.E., M.A., M.I.E.E.

Deputy Secretary

F. JERVIS SMITH, M.I.E.E.

Assistant Secretary

F. C. HARRIS.

Editor-in-Chief

G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E.

The Institution is not, as a body, responsible for the opinions expressed by individual authors or speakers.
An example of the preferred form of bibliographical references will be found beneath the list of contents.

THE PROCEEDINGS OF
THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 103. PART C. No. 3. MARCH 1956

621.316.89.029.6

Monograph No. 132 R
May 1955

THE THEORY AND DESIGN OF COAXIAL RESISTOR MOUNTS FOR THE FREQUENCY BAND 0–4 000 Mc/s

By I. A. HARRIS, Associate Member.

(The paper was first received 15th May, 1954, and in revised form 27th January, 1955. It was published as an INSTITUTION MONOGRAPH *in May, 1955.)*

SUMMARY

An account is given of the theoretical basis for the design of coaxial resistors that retain their d.c. resistance, without appreciable reactance, at all frequencies at which coaxial systems are normally used. The main restriction on physical size is governed by the avoidance of supplementary modes of propagation. The design employs a uniform cylindrical film resistor with a critically dimensioned outer conductor, the profile of which has the form of a tractrix. Lead-in cones are designed to avoid discontinuity at the connections with the resistor and the outer conductor. Experimental results show an impedance within 1% of the d.c. resistance, with an extremely small phase angle, at all frequencies up to the highest measured, namely 3 450 Mc/s.

(1) INTRODUCTION

The realization of constant resistance without reactance over the whole range of frequencies at which coaxial lines are used is a problem encountered in the design of coaxial loads, dissipative attenuators or "standard" r.f. resistors for impedance bridges or other measuring apparatus. In the best-known attempt to solve the problem, a cylindrical resistive film is used as the inner conductor of a short, uniform coaxial line which is otherwise loss-free. By applying the ordinary theory of uniform transmission lines to the problem, it has been shown[1] that with one end of such a coaxial resistor short-circuited the impedance at the other end can be made resistive and substantially independent of frequency by suitable choice of the ratio of inner- and outer-conductor diameters. Under this condition, the resistance does not vary with frequency by more than 1%, provided that the wavelength corresponding to the highest frequency applied is not less than about 30 times the resistor length. By employing a different diameter ratio, the restriction on the length of the resistor may be relaxed at the expense of introducing a small susceptance, although this may be compensated over a wide frequency band either by undercutting the inner lead-in conductor for a short length or by other and better means.[2] Even with this arrangement, resistors for frequencies up to 3 000 Mc/s are restricted in length to about 1 cm, which severely limits their power rating. Any advantage to be gained by using a liquid dielectric coolant is offset by the higher permittivity, which reduces the wavelength and thus the permissible length of resistor.

A more successful mode of attack on the problem is based on a new interpretation of the well-known rule that if a uniform, lossless transmission line be terminated in a resistance equal to the (resistive) characteristic impedance of the line, the input impedance of the system is equal to the terminating resistance. This condition is independent of the length of the line; in particular, it is true when the line is extremely short. Consider a cylindrical film resistor of resistance R divided into four equal sections, with a short piece of loss-free coaxial line interposed between each pair of sections, as shown in Fig. 1. At the cross-section A the resistance seen to the right is R, and the characteristic impedance of the coaxial line leading-in to A is made equal to R. At the cross-section B the resistance seen to the right is $\frac{3}{4}R$, and the characteristic impedance of the coaxial line leading-in to B is made equal to $\frac{3}{4}R$. Likewise at C and D the characteristic impedances of the respective lead-in coaxial lines are made equal to $\frac{1}{2}R$ and $\frac{1}{4}R$. If, in Fig. 1, the length of each inter-section loss-free line be made infinitesimal, the profile of the outer conductor is then determined at the four cross-sections A, B, C and D. By increasing indefinitely the number of sections into which the resistor is divided, the continuous profile of the outer conductor is determined. It is readily seen that, if R_0 is the resistance per unit length and w is a distance from the short-circuited end, the characteristic impedance of a loss-free coaxial line with radii equal to those of the resistor and the outer con-

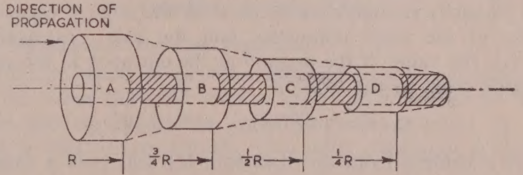


Fig. 1.—Cylindrical film resistor divided into four equal parts, with sections of loss-free coaxial line interposed.

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Harris is in the Laboratories of the Aeronautical Inspection Directorate, Ministry of Supply.

ductor at the cross-section at w must equal wR_0 if reflectionless propagation along the resistor is to be obtained. The characteristic impedance Z_c of a loss-free coaxial line is given by the well-known relationship

$$Z_c = (\zeta/2\pi)\log_\varepsilon(r/a) \text{ ohms} \quad . \quad . \quad . \quad . \quad (1)$$

in which r and a are the radii of the outer and inner conductors respectively and $\zeta = (\mu/\epsilon)^{1/2}$ is the wave impedance, in ohms, of a plane wave, μ being the absolute permeability (henry/metre) and ϵ the absolute permittivity (farad/metre) of the medium. Equating Z_c to wR_0 gives the equation of the profile of the outer conductor,[3] namely

$$r = a\exp(2\pi R_0 w/\zeta) \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The validity of eqn. (2) depends upon the supposition that the wavefront is always plane and normal to the axis, so that the system may legitimately be divided into cylindrical "slices" which do not intersect a wavefront at any instant. In so far as the actual field approximates to this requirement, a resistor mount designed according to eqn. (2) should be perfect at all frequencies at which only the principal wave is propagated in the system. The longitudinal section of a resistor mount consistent with eqn. (2) is shown in Fig. 2. In accordance with the laws of the electromagnetic field, the direction of the electric field at the outer conductor (assumed perfect) is normal to the conductor surface, while the electric-field direction at the resistor surface makes a small angle with the normal to the surface. The resulting form of the electric field (and with it the wavefront) is indicated by the set of broken lines in Fig. 2. It is seen that the departure from planarity is largely conditioned by the angle θ between the section of the outer conductor and the axis. According to eqn. (2), the value of the angle θ at the entrance to the resistor mount is given by the relation

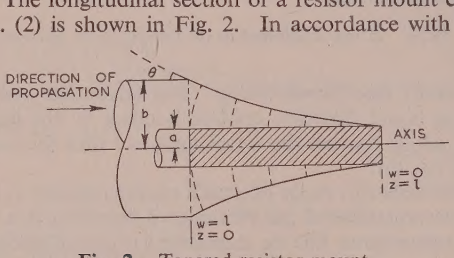


Fig. 2.—Tapered resistor mount.

The natural direction of the electric field is indicated by the broken lines.

$$\tan\theta = (dr/dw)_{w=1} = 2\pi R_0 b/\zeta \quad . \quad . \quad . \quad . \quad (3)$$

and it is evident from this and eqn. (1) that only a long thin resistor ensures the validity of eqn. (2). In general, a resistor mount designed according to eqn. (2) will be imperfect, and the greater the resistance per unit length and the greater the total resistance, the more imperfect the mount will be.

To achieve the aim stated at the beginning of the Section, it is evident that a design is required which, although based on principles similar to the foregoing, also takes into account the departure from planarity of the guided principal wave in the resistor mount. Accordingly, the theory of a special class of non-uniform transmission systems must be developed from the basic electromagnetic equations and then be applied to the design of a tapered resistor mount.

(2) THEORETICAL ASPECTS OF SOME NON-UNIFORM COAXIAL TRANSMISSION SYSTEMS

(2.1) Approximations inherent in the Transmission-Line Equations for Uniform Resistive Coaxial Lines

The ordinary transmission-line equations are based on the following differential equations for a uniform dielectric bounded by a pair of loss-free conductors:

$$\left.\begin{aligned} -\partial I/\partial z &= G_0 V + C_0 \partial V/\partial t \\ -\partial V/\partial z &= L_0 \partial I/\partial t \end{aligned}\right\} \quad . \quad . \quad . \quad (4)$$

in which G_0, C_0, L_0 are the distributed shunt conductance, shunt capacitance and series inductance per unit length in the z-direction, respectively. These equations follow rigorously* from the form of Maxwell's equations in which the electric field is radial in a cylindrical co-ordinate system and the wavefront is therefore planar. In a system with appreciable conductor resistance, however, there is necessarily a component of electric field tangential to the conductor surface in the z-direction, so that the resultant electric field cannot be wholly radial and the wavefront cannot be planar. But, as Heaviside pointed out,[4] so long as the resistance R_0 per unit length is small, it is usually sufficient to add the term $R_0 I$ to the right-hand side of the second of eqns. (4) and ignore the slight distortion of the wavefront. This approximation is inherent in the familiar transmission-line equations, and is of special interest in the present work, in which coaxial lines with cylindrical resistive-film inner conductors are considered. For this reason, a rigorous solution of the problem of a uniform coaxial line with a perfect outer conductor and a resistive-film inner conductor is given in Section 7.1. For the range of values likely to be encountered in practice, the results show that the ordinary transmission-line equations with L_0, C_0 and R_0 calculated from the simple (radio frequency) formulae are a first approximation. To a second approximation the effects of the departure from planarity of the wave and of penetration of the field into the enclosure formed by the cylindrical film have to be taken into account. It is shown in Section 7.1 that this may be done simply by modifying the value of L_0 given by the well-known formula for perfect conductors.

(2.2) Limitations in the Application of Transmission-Line Equations to Non-Uniform Resistive Coaxial Lines

To investigate how far transmission line equations may be applied to non-uniform lines, it has to be determined under what conditions the "general principal wave," which conforms with the general orthogonal co-ordinate system conditioned by the guiding conductors, exists. The concept "general principal wave" is an extension embracing the concepts "plane wave," "spherical wave," and any other form that satisfies certain requirements to be set out later. Thus, between a uniform loss-free cylindrical coaxial pair of conductors, the wavefront is planar, while between a coaxial pair of loss-free conical conductors with a common apex the wavefront is spherical; in both examples the wavefront naturally conforms with the orthogonal co-ordinate system conditioned by the guiding conductors. By analogy, the wavefront of the general principal wave conforms with the general orthogonal co-ordinate system conditioned by the (coaxial) conductor pair.

Orthogonal curvilinear co-ordinates u_1, u_2, u_3 are used, for which an arbitrary line element ds is given by the differential form

$$ds^2 = h_1^2 du_1^2 + h_2^2 du_2^2 + h_3^2 du_3^2$$

where, relative to a Cartesian co-ordinate system x, y, z

$$h_1 = +[(\partial x/\partial u_1)^2 + (\partial y/\partial u_1)^2 + (\partial z/\partial u_1)^2]^{1/2} \quad . \quad (5)$$

with corresponding relations for h_2 and h_3. All electromagnetic quantities will be expressed in rationalized M.K.S. units. It is assumed that each surface u_3 = (a constant) constitutes a wavefront in which E and H, the electric field intensity and the

* At least for coaxial systems from which there is no radiation.

magnetic field intensity respectively, are tangential and mutually perpendicular. The components E_1 and H_2 alone exist in a right-handed system of co-ordinates u_1, u_2, u_3. For the field between the conductor pair, the form of the electromagnetic field equations is

$$\left.\begin{aligned}\operatorname{curl} \boldsymbol{H} &= \epsilon \partial \boldsymbol{E}/\partial t \\ \operatorname{curl} \boldsymbol{E} &= -\mu \partial \boldsymbol{H}/\partial t\end{aligned}\right\} \quad \ldots\ldots \quad (6)$$

in which ϵ and μ are constants for a homogeneous isotropic medium. If these equations be expressed in curvilinear co-ordinates and the above restrictions on $\boldsymbol{E}$ and $\boldsymbol{H}$ are imposed, two significant equations follow, namely

$$-1/(h_2h_3).\partial(h_2H_2)/\partial u_3 = \epsilon \partial E_1/\partial t \quad \ldots \quad (7)$$

$$1/(h_1h_3).\partial(h_1E_1)/\partial u_3 = -\mu \partial H_2/\partial t \quad \ldots \quad (8)$$

Differentiation of eqn. (7) by u_3 after multiplication by h_1 and substitution for h_1E_1 from eqn. (8) leads to

$$h_2/(h_1h_3).\partial[h_1/(h_2h_3).\partial(h_2H_2)/\partial u_3]/\partial u_3 = \mu\epsilon\partial^2(h_2H_2)/\partial t^2 \quad (9)$$

and the corresponding result from eqn. (8) is

$$h_1/(h_2h_3).\partial[h_2/(h_1h_3).\partial(h_1E_1)/\partial u_3]/\partial u_3 = \mu\epsilon\partial^2(h_1E_1)/\partial t^2 \quad (10)$$

Inspection of eqns. (9) and (10) reveals that, if both h_3 and h_1/h_2 are independent of u_3, they may be written

$$\left.\begin{aligned}1/h_3^2.\partial^2(h_2H_2)/\partial u_3^2 &= \mu\epsilon\partial^2(h_2H_2)/\partial t^2 \\ 1/h_3^2.\partial^2(h_1E_1)/\partial u_3^2 &= \mu\epsilon\partial^2(h_1E_1)/\partial t^2\end{aligned}\right\} \quad \ldots \quad (11)$$

which have the well-known wave solutions

$$\left.\begin{aligned}h_2H_2 &= C_1\mathrm{f}(t - h_3u_3/v) + C_{-1}\mathrm{F}(t + h_3u_3/v) \\ h_1E_1 &= C_2\mathrm{f}(t - h_3u_3/v) + C_{-2}\mathrm{F}(t + h_3u_3/v)\end{aligned}\right\} \quad . \quad (12)$$

in which $v = (\mu\epsilon)^{-1/2}$ is constant and h_3 is necessarily independent of u_3. The functions f and F are the same for H_2 and E_1, because, as well as being solutions of eqns. (11), they have also to satisfy eqns. (7) and (8) separately.

The conditions that h_3 and h_1/h_2 be independent of u_3 are similar to the restrictions imposed initially by Bromwich in his general solution of Maxwell's equations,[5] except that Bromwich imposed the more severe restriction $h_3 = 1$ on the co-ordinate system. In seeking the physical meaning of these conditions governing the strict validity of the concept "general principal wave," it is instructive to examine the approximate solution of eqns. (7) and (8) when h_3 varies slowly with u_3. There is a similar problem in physical optics: the approximate solution of a wave equation in which the refractive index $(\mu\epsilon/\mu_0\epsilon_0)^{1/2}$ varies from place to place. The result shows that, if the refractive index varies little over the distance of a wavelength, the propagation is essentially that of geometrical optics. Here, an approximate solution of eqns. (7) and (8) for a wave travelling in the positive u_3 direction is sought in the form of eqns. (12), in which the former constants C_1, C_2 and h_3/v may now vary slowly with u_3. Substituting the trial solutions in eqns. (7) and (8) separately gives

$$C_1 = \text{constant},\ C_2 = \text{constant} \quad \ldots \quad (13)$$

$$C_1\Lambda/h_2v = \epsilon C_2/h_1 \quad \ldots\ldots \quad (14)$$

$$C_2\Lambda/h_1v = \mu C_1/h_2 \quad \ldots\ldots \quad (15)$$

in which

$$\Lambda = 1 + u_3v/h_3.\partial(h_3/v)/\partial u_3 \quad \ldots\ldots \quad (16)$$

From eqns. (14) and (15) there follow

$$H_2 = \epsilon vE_1\Lambda^{-1} \quad \ldots\ldots \quad (17)$$

$$E_1 = \mu vH_2\Lambda^{-1} \quad \ldots\ldots \quad (18)$$

from which the phase velocity v is found to be

$$v = (\mu\epsilon)^{-1/2}\Lambda \quad \ldots\ldots \quad (19)$$

In eqn. (16) the condition for slow variation of h_3 with u_3 is expressed by $u_3v/h_3.\partial(h_3/v)/\partial u_3 \ll 1$, so that $\Lambda \to 1$. The variable term in Λ contains h_3, which may be a function of u_1 and u_2, co-ordinates of position on a wavefront, so that if h_3 varies at all with u_3, the velocity v will in general be different at different points on the wavefront. This can only mean that there is a slow bending of the wavefront as the wave progresses, such as would occur in traversing a slight bend in a coaxial line. In this example, it is clear that the wave velocity must be greater on the outside of a bend than on the inside. The requirement that h_3 be independent of u_3 therefore means that there must be no bending of the wavefront from the shape conditioned by uniform propagation, in order to conform with the bounding conductors which govern the co-ordinate system. The other condition, namely that h_1/h_2 be independent of u_3, means that there must be no rotation of the plane(s) of polarization as the wave progresses.

As well as considering co-ordinate systems which strictly satisfy these conditions, systems in which h_3 varies very slowly with u_3 will also be considered. For such systems, eqns. (17) and (18) yield the following important relation:

$$E_1 = \zeta H_2 \quad \ldots\ldots \quad (20)$$

in which $(\mu/\epsilon)^{1/2} = \zeta$ is the wave impedance of the medium for principal waves. Eqn. (20), well-known for plane waves, is therefore valid for general principal waves which either satisfy the conditions rigorously or satisfy them only approximately in so far as h_3 varies slowly with u_3.

The possibility of describing propagation in a system by simple wave equations such as eqn. (11), and conformity with eqn. (20), are also conditions for the applicability of "distributed circuit constant" or "transmission-line" equations to the study of the system. Thus transmission-line equations are approximately valid when the conditions are relaxed to permit h_3 to vary slowly with u_3.

(2.3) Transmission-Line Equations for Non-Uniform Resistive or Loss-Free Coaxial Lines

(2.3.1) Derivation of the Line Equations.

In the derivation of the representation of a general principal wave (in either the strict or the approximate sense) by transmission-line equations, the class of system typified by that shown in Fig. 3 is considered. The inner conductor is either a uniform resistive film of resistance R_s per square or is a good conductor, while the outer conductor has an arbitrary taper and is stipulated to be a perfect conductor.

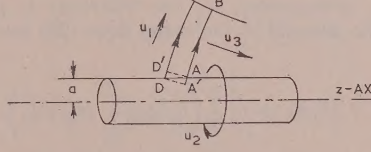


Fig. 3.—The orthogonal co-ordinate system applied to a non-uniform coaxial system with a resistive inner conductor.

Let the second circuital law of the electromagnetic field,

$$\oint \boldsymbol{E}dl = -\mu\frac{\partial}{\partial t}\iint \boldsymbol{H}dS \quad \ldots \quad (21)$$

(where dl is an element of the contour around the surface S) be applied to the area ABCDA in Fig. 3. The positive direction is taken in this order. The line integral of $\boldsymbol{E}$ along a definite curve between two points defines the voltage V between the points, even though this is not necessarily the electrostatic p.d. when the field varies in time, as is evident from eqn. (21). The path BC in Fig. 3 contributes nothing to the line integral, because the direction of $\boldsymbol{E}$ is everywhere normal to u_3, but the path DA contributes on account of the p.d. along the resistor. Disregarding the internal reactance of the resistor and the reactance arising from the field penetrating the resistive film, the p.d. along the resistor is $IR_0\delta z$, where R_0 $(=R_s/2\pi a)$ is the resistance per unit length and z is distance along the resistor surface. This contribution is bound up with the fact that the electric-field vector meets the resistor surface at an angle to the normal. which is denoted by ψ. In Fig. 4, the path DA is vectorially equivalent to the path A′A in the line integral of eqn. (21), because, vectorially, DA = DA′ + A′A and there is no component of $\boldsymbol{E}$ along DA′. Then $IR_0\delta z = E_{1(a)} \sin \psi\delta z$ in which the subscript (a) refers to the resistor surface. It follows that

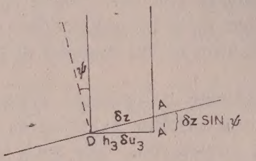


Fig. 4.—Enlargement of detail shown in Fig. 3.

$$IR_0 = E_{1(a)} \sin \psi \quad . \quad . \quad . \quad . \quad . \quad (22)$$

in which the subscript (a) refers to the resistor surface. The current I in the resistor is given by the line integral of H around the resistor, and

$$I = H_{2(a)} 2\pi a \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

Eqns. (22) and (23), together with eqn. (20), lead to the important result

$$\sin \psi = R_s/\zeta \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

Applying the usual procedure to the path ABCD, and adding the contribution $IR_0\delta z$ for the part DA, gives

$$\oint_{ABCDA} \boldsymbol{E}d\boldsymbol{l} = \delta V + IR_0\delta z \quad . \quad . \quad . \quad . \quad (25)$$

in which $\delta V = V_{AB} - V_{DC}$, the voltage increment at AB over the voltage at DC. The surface integral in eqn. (21) is taken over the area ABCDA. (Whether ABCD′A or A′BCDA′ be taken is of no consequence, for the difference is of an order smaller than the integral.) On using eqn. (20) and eqn. (25) there follows

$$\delta V + IR_0\delta z = -\frac{\partial}{\partial t}\left(h_3\delta u_3\mu \int_D^C H_2 h_1 du_1\right) \quad . \quad . \quad (26)$$

In the derivation of eqn. (26) it is to be noted that δz is related to δu_3 by the expression $\delta z = \delta u_3[dz/du_3]_{(a)}$, or on reference to Fig. 4, $\delta z = h_3\delta u_3/\cos \psi$. The expression in brackets in eqn. (26) is the magnetic flux linking the contour ABCDA which, according to the definition of self-inductance L_0 per "unit length," may be written $L_0 h_3\delta u_3 I$. The self-inductance per unit length in u_3 is therefore

$$L_0 = \frac{\mu}{I}\int_D^C H_2 h_1 du_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

where I is given by the line integral of $\boldsymbol{H}$ around the inner conductor. Substituting eqn. (27) for the integral in eqn. (26), dividing by $h_3\delta u_3$ and passing to the limit $\delta u_3 \to 0$ gives

$$\frac{1}{h_3}\frac{\partial V}{\partial u_3} = L_0\frac{\partial I}{\partial t} + \left(\frac{1}{h_3}\frac{dz}{du_3}\right)_{(a)} R_0 I \quad . \quad . \quad . \quad (28)$$

This corresponds to the second transmission-line equation [cf. eqn. (4)] and refers to propagation in the natural u_3 direction. For those cases in which h_3 varies slowly with u_3, it is more useful to refer the propagation to the z-direction along the resistor surface. The equation then becomes

$$-\partial V/\partial z = L_0'\partial I/\partial t + R_0 I \quad . \quad . \quad . \quad . \quad (29)$$

in which $L_0' = L_0 \cos \psi$.

Next, let the first circuital law of the electromagnetic field

$$\oint \boldsymbol{H}d\boldsymbol{s} = \epsilon\frac{\partial}{\partial t}\iint \boldsymbol{E}d\boldsymbol{S} \quad . \quad . \quad . \quad . \quad (30)$$

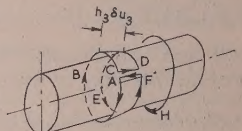


Fig. 5.—Contour for the line integral of H.

(where ds is an element of the contour around the surface S) be applied to the contour ABCDEF shown in Fig. 5. On applying the usual procedure there is obtained

$$-\delta I = \frac{\partial}{\partial t}\left(h_3\delta u_3 \oint_{ABCA} \epsilon E_1 h_2 du_2\right) \quad . \quad . \quad . \quad (31)$$

in which δI is the increment in the current at F over that at A, and the expression in brackets is the electric flux threading the closed curve. This flux is equal to the equivalent surface charge on the adjacent conducting surface. If V be the voltage between the conductors, this charge is $h_3\delta u_3 C_0 V$, where C_0 is the capacitance per unit length in the u_3 co-ordinate. Then

$$C_0 = \frac{\epsilon}{V}\oint E_1 h_2 du_2 \quad . \quad . \quad . \quad . \quad . \quad (32)$$

in which V is given by the line integral

$$V = \int_a^b E_1 h_1 du_1 \quad . \quad . \quad . \quad . \quad . \quad (33)$$

taken along a curve in the surface u_3 = a constant. Substituting eqn. (33) for the integral in eqn. (31), dividing by $h_3\delta u_3$ and proceeding to the limit $\delta u_3 \to 0$ gives

$$-1/h_3 . \partial I/\partial u_3 = C_0\partial V/\partial t \quad . \quad . \quad . \quad . \quad (34)$$

This corresponds to the first transmission-line equation [cf. eqn. (4)] and refers to propagation in the u_3-direction. As with eqn. (29), the equation referring the propagation to the z-direction is

$$-\partial I/\partial z = C_0'\partial V/\partial t \quad . \quad . \quad . \quad . \quad . \quad (35)$$

in which $C_0 = C_0 \cos \psi$.

Eqns. (28) and (34) are most useful in those problems in which the conditions of validity of a general principal wave are rigorously satisfied, while eqns. (29) and (35) are most useful in those problems in which h_3 varies slowly with u_3, and the transmission-line equations are therefore approximate. As would be expected for a wave referred to a direction z at an angle ψ to the natural direction of propagation, the phase velocity in the z-direction is $(L_0'C_0'/L_0C_0)^{-1/2} = 1/\cos\psi$ times that in the natural direction. Also, the relation $L_0'/C_0' = L_0/C_0$ obtains.

(2.3.2) Characteristic Impedance in Non-Uniform Systems.

Wave impedance ζ is a property of the wave and the medium. For a general principal wave in a loss-free medium, $\zeta = (\mu/\epsilon)^{1/2}$ and is resistive [eqn. (20)], while in a lossy medium the wave impedance is complex. In this work, losses in the dielectric are assumed to be negligible compared with other losses, and the complex form of ζ will not be required.

Characteristic impedance, Z_c, at any section of a coaxial system is a property of the wave impedance and of the proportions of the guiding conductors, its value being given by the product of the wave impedance and a numerical factor determined by the geometry of the system. For a guided principal wave, Z_c is the ratio of voltage to current at the wavefront in question, determined by the value of u_3. According to eqns. (33) and (12) with F ≡ 0 and the relation $I = \oint Hds$, there results the general equation

$$Z_c(u_3) = \left(\frac{V}{I}\right)_{u_3} = \frac{\int_a^b E_1h_1du_1}{\oint H_2h_2du_2} = \frac{\zeta}{2\pi}\int_a^b \frac{h_1}{h_2}du_1 \quad . \quad (36)$$

in which a and b refer to the inner and outer conductors respectively. When the inner conductor is resistive, the value of Z_c thus defined is better termed the "wavefront characteristic resistance" to distinguish it from the complex characteristic impedance of a uniform resistive transmission line as used in ordinary line theory. An application of eqns. (27), (33) and (36) shows that the familiar relation $Z_c = (L_0/C_0)^{1/2}$ applies to all valid general systems.

(2.4) Reflectionless Propagation in Resistive Coaxial Systems

(2.4.1) Comparison between Resistive Systems with Uniform and Tapered Outer Conductors.

The field of a uniform coaxial system with a resistive inner conductor and a perfect outer conductor is investigated in Section 7.1, and the resulting form of the electric field is shown in section in Fig. 6. The wavefront continually bends towards the resistor as the wave progresses, and the energy equivalent to those (fictitious) parts of the field shown shaded in Fig. 6 is converted into heat in the resistor. According to the conclusions of Section 2.2, this bending of the direction of propagation makes the application of transmission-line formulae valid only as an approximation. Both $\boldsymbol{E}$ and $\boldsymbol{H}$ are attenuated as the wave progresses, by the "thinning out" of the electric and magnetic flux as a result of the bending and dissipation in the resistor. A reactive component of the characteristic impedance associated with the approximate transmission-line equation also appears. The power dissipated in unit length of the resistor decreases exponentially along the system, with the result that nearly all the power is dissipated near one end of the resistor.

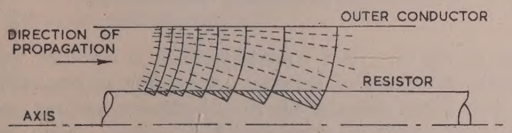


Fig. 6.—Representation of dissipation in the resistive inner conductor of a uniform coaxial line.

The Poynting flux of energy is shown by the dotted lines. An equal rate of dissipation occurs in each shaded triangle, the area of the triangle representing the equivalent amount of field at the place in question.

The form of the electric field of a system with a uniform, resistive inner conductor and a tapered perfect outer conductor, similar to that shown in Fig. 2, is shown in Fig. 7. The nature of the flow of energy from the electromagnetic field to the resistor may be seen by a simple application of one form of Poynting's theorem. If, in Fig. 7, energy crosses the surface AA′ at the rate P_A (watts), energy is dissipated in the resistor AB at the rate δP_R and energy leaves the surface BB′ at the rate P_B, and if there is no storage within the enclosure AA′BB′, the continuous application of the law of conservation of energy requires that

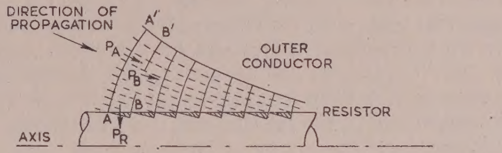


Fig. 7.—Representation of dissipation in the resistive inner conductor of a tapered coaxial line, in which the electric and magnetic flux densities remain constant while the wavefront area is progressively reduced.

Compare with Fig. 6.

$$P_A = P_B + \delta P_R \quad . \quad . \quad . \quad . \quad . \quad (37)$$

For quantities that vary sinusoidally with time, Poynting's theorem shows that the time mean of the energy flux (power) $\boldsymbol{P}$ leaving an enclosure is the integral

$$\boldsymbol{P} = \iint \overline{\boldsymbol{E} \times \boldsymbol{H}} d\boldsymbol{S} \text{ watts} \quad . \quad . \quad . \quad . \quad (38)$$

taken over the surface.* The instantaneous energy flux varies sinusoidally between 0 and 2$\boldsymbol{P}$ twice every cycle. In the case of a general principal wave, the time mean power per unit area passing any part of the surface is given by $\boldsymbol{P}_0 = \overline{\boldsymbol{E} \times \boldsymbol{H}}$ at the place in question. On applying eqn. (38) to the wave of eqns. (12), with f ≡ sin and F ≡ 0, the mean power crossing AA′ from left to right is

$$P_A = \int_A^{A'} du_1 \oint du_2h_1h_2\overline{E_1H_2}$$

and since $E_1 = \zeta H_2$ and $I = \oint H_2h_2du_2$

$$P_A = \frac{\zeta\hat{I}^2}{4\pi}\int_A^{A'} \frac{h_1}{h_2}du_1 \quad . \quad . \quad . \quad . \quad . \quad (39)$$

or, by eqn. (36),

$$P_A = \tfrac{1}{2}\hat{I}^2(Z_c)_A \quad . \quad . \quad . \quad . \quad . \quad (40)$$

where $\hat{I}$ is the peak value of I. On stipulating that the current amplitude I is constant along the resistor, such as would produce uniform dissipation in a uniform resistor, the mean power crossing BB′ from left to right is similarly

$$P_B = \tfrac{1}{2}\hat{I}^2(Z_c)_B \quad . \quad . \quad . \quad . \quad . \quad (41)$$

* $\boldsymbol{E}$, $\boldsymbol{H}$, $\boldsymbol{I}$ are here instantaneous values of sinusoidally varying quantities. The bar denotes the time mean.

The power dissipated in the resistor between A and B is $\delta P_R = \frac{1}{2}\hat{I}^2 R_0 \delta z$, so that with eqns. (40), (41) and (37)

$$(Z_c)_A - (Z_c)_B = R_0\delta z \quad . \quad . \quad . \quad . \quad (42)$$

Thus, with the same current I at A and at B, the difference between the wavefront characteristic resistances at A and B equals the resistance between A and B. Because $I_A = I_B$ in amplitude, $(H_2)_A = (H_2)_B$ in amplitude, and, with the relationship $E_1 = \zeta H_2$, $(E_1)_A = (E_1)_B$ in amplitude also. Therefore both $\boldsymbol{E}$ and $\boldsymbol{H}$ remain unchanged in amplitude during such reflectionless propagation, the difference between $(Z_c)_A$ and $(Z_c)_B$ required by eqn. (42) being achieved by reduction of the area of the wavefront as it progresses. This can be seen from inspection of eqn. (36), in which E_1 and H_2 are constant in amplitude while the value of $Z_c(u_3)$ can be made to vary with u_3 by varying the path length a, b in the line integral, i.e. by tapering the outer conductor.

To sum up, the attenuation of the system shown in Fig. 7 is by reduction of the wavefront area as the wave progresses, the energy of that part of the field which is thus removed being converted into heat in the resistor. By contrast with the uniform system shown in Fig. 6, there is less "bending" of the wavefront during propagation and the dissipation is uniformly distributed along the resistor. Eqn. (42) expresses the condition for reflectionless propagation which was derived in the Introduction by a more elementary method.

(2.4.2) Application of Transmission-Line Equations to the Problem.

Let the time variation of the field quantities be sinusoidal, i.e. proportional to the real part of exp $(j\omega t)$. Because coaxial systems with a uniform cylinder of resistive material as the inner conductor in general admit only approximate solution by transmission-line equations, the approximate equations (29) and (35) are used as a starting-point.* There is obtained:

$$-dV/dz = (j\omega L_0' + R_0)I \quad . \quad . \quad . \quad . \quad (43)$$

$$-dI/dz = j\omega C_0' V \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

Although these equations are superficially the same as the ordinary uniform line equations, they differ in that C_0' and L_0' may now be functions of z. This dependence on z occurs through the upper limit of the line integral in eqn. (27) and similarly in eqn. (33). Differentiation of eqn. (44) by z gives

$$-d^2I/dz^2 = j\omega C_0'(dV/dz) + j\omega(dC_0'/dz)V$$

and substitution for dV/dz from eqn. (43) and for V from eqn. (44) leads to the differential equation

$$\frac{d^2I}{dz^2} - \frac{1}{C_0'}\frac{dC_0'}{dz}\frac{dI}{dz} + (\omega^2 L_0' C_0' - j\omega C_0' R_0)I = 0 \quad . \quad (45)$$

Let the independent variable be changed to w, where $w = l - z$, l being the length of the resistor (Fig. 2). The variation of L_0' and C_0' with z must be such that eqn. (42) is satisfied, and on noting that $Z_c = 0$ at $w = 0$, we have $(L_0'/C_0')^{1/2} = Z_c = wR_0$. Uniformity of velocity of propagation along the resistor requires $L_0'C_0' =$ a constant ($= K^2$ say). From this, two relationships follow, namely

$$L_0' = wR_0K \quad \text{and} \quad C_0' = K/(wR_0)$$

On writing $\beta = \omega K$ and making use of the two preceding relations, eqn. (45) becomes

$$d^2I/dw^2 + (1/w)dI/dw + (\beta^2 - j\beta/w)I = 0$$

This differential equation may be solved by substituting* $\varepsilon^{-j\beta w}T(w)$ for I in it, leading to

$$wd^2T/dw^2 + (1 - 2j\beta w)dT/dw - 2j\beta T = 0$$

which, in turn, may be solved as a power series in w according to the method of Frobenius.[6] The solution is

$$T = A_1\varepsilon^{2j\beta w} + A_2[\varepsilon^{2j\beta w}\log_\varepsilon w - 2j\beta w + \tfrac{3}{4}(2\beta w)^2 - \ldots]$$

in which A_1 and A_2 are two arbitrary constants. In the physical problem (Fig. 2) the point $w = 0$ is included in the domain of the solution, but because $\log_\varepsilon w$ becomes infinite at this point, the constant A_2 must vanish in order to keep T finite. Therefore $T = A_1\varepsilon^{2j\beta w}$, whence $I = A_1\varepsilon^{j\beta w}$, or in terms of z, including the time factor which has been suppressed from eqn. (43) onwards,

$$I = A\varepsilon^{j(\omega t - \beta z)} \quad . \quad . \quad . \quad . \quad . \quad (46)$$

This result shows that I is constant in amplitude. The voltage V between the ends of a line of electric force associated with the position z along the resistor follows from eqns. (46) and (44), and is expressed as

$$V = A(l - z)R_0\varepsilon^{(\omega t - \beta z)} \quad . \quad . \quad . \quad . \quad (47)$$

The impedance at any wavefront, V/I, is clearly given by $(l - z)R_0$ and is resistive.

(3) THE DESIGN OF A RESISTOR MOUNT

(3.1) Formulation of the Outer Conductor Profile

Given a uniform resistive film in the form of a right circular cylinder, the determination of the profile of the outer conductor is a matter of mathematical trial and error rather than a deductive process. Having assumed a given configuration, its conformity with the simple condition for reflectionless propagation given in the preceding Section is then checked by the methods developed in Sections 2.2 and 2.3.

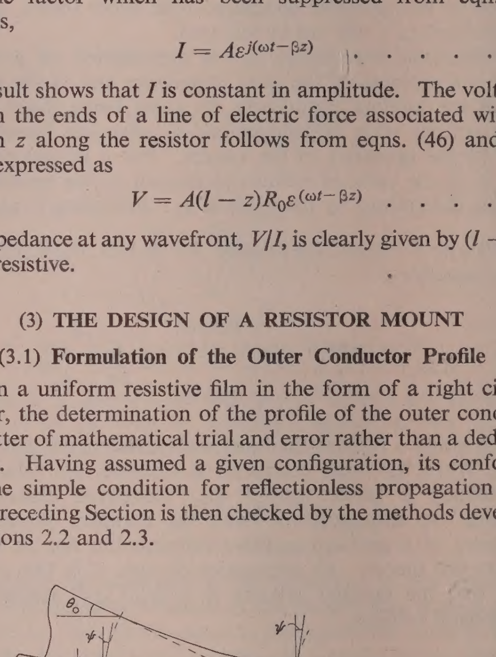


Fig. 8.—Nomenclature and geometrical arrangement for resistor mount with coaxial cone input terminals.

Refer to Fig. 8. As a starting point it is assumed that the wavefronts are parts of spheres. According to eqn. (24), the angle ψ between the normal to the resistor surface and the electric field vector is given by $\sin\psi = R_s/\zeta$. Then, with a uniform resistor, ψ is constant at all points along the resistor, and on inspecting Fig. 8 it is seen that this constrains all the part-spheres to be of the same radius $\rho_0 (= a/\sin\psi = a\zeta/R_s)$, the locus of their centres being the axis of the coaxial system. The orthogonal co-ordinate system resulting from this is derived in Section 7.2 and is illustrated in Fig. 12. It is shown that h_1/h_2 is constant, and provided that $R_s/\zeta \ll 1$, h_3 varies slowly with u_3, so that the conditions for the validity of a principal wave are approximately satisfied. It remains to show that the condition

* For a conical inner conductor of uniform surface resistance R_s, and of semi-angle ψ given by eqn. (24), a rigorous solution exists, eqns. (28) and (34) being most suitable for this problem. The co-ordinates are cylindrical and the wavefront is planar. In practice, however, the rigour is upset by the effect of the field penetrating the resistive film, while the utility of the design is adversely affected by the uneven distribution of power dissipation in the resistor.

* The substitution $\varepsilon^{j\beta w}T(w)$ would equally serve, a different value of T being obtained, leading to the same value of I at the end of the calculation.

for reflectionless propagation is satisfied. Eqn. (97) [Section 7.2] may be written

$$u_3 - u_1 = \rho_0 \log_\varepsilon \left[\frac{\tan(\theta/2)}{\tan(\psi/2)}\right] \quad . \quad . \quad . \quad (48)$$

in which θ is a function of u_1 and u_3 defined by this relationship, and ψ is the constant angle already defined. The wavefront characteristic resistance is given by eqn. (36), with $h_1/h_2 = 1/\rho_0$ as shown by eqns. (101), as

$$Z_c(u_3) = \frac{\zeta}{2\pi}\int_a^b \frac{du_1}{\rho_0} = \frac{\zeta}{2\pi}\left[\frac{(u_1)_b - (u_1)_a}{\rho_0}\right]$$

in which a and b refer to the two ends of a curve u_3 = a constant (Fig. 12). On applying eqn. (48), we obtain

$$Z_c(u_3) = \frac{\zeta}{2\pi}\log_\varepsilon\left[\frac{\tan(\theta_0/2)}{\tan(\psi/2)}\right] \quad . \quad . \quad . \quad (49)$$

in which θ_0 is the angle which the curve $u_1 = 0$ makes with the axis at the point where it cuts the curve u_3 = a constant. Again, on using eqn. (24) there results $R_0 w = (\zeta/2\pi)w/\rho_0$.

When associating a result with a curve u_3 = a constant, u_3 may be equated to w, so that from eqn. (48) there is obtained

$$R_0 w = \frac{\zeta}{2\pi}\log_\varepsilon\left[\frac{\tan(\theta_0/2)}{\tan(\psi/2)}\right] \quad . \quad . \quad . \quad (50)$$

On comparing eqns. (49) and (50) it is seen that the condition for reflectionless propagation is satisfied within the limits of the approximation made in assuming the existence of a principal wave.

It is shown in Section 7.2 that the radius r of the outer conductor is given as a function of position w along the axis (Fig. 8) by the formula[7]

$$r \simeq \frac{a(1 + s^2/4)\exp[(w + \Delta w)/\rho_0]}{1 + (s^2/4)\exp[2(w + \Delta w)/\rho_0]} \quad . \quad . \quad (51)$$

where
$$\left.\begin{aligned}\Delta w &\simeq (as/2)[\exp(2w/\rho_0) - 1]\\ s &= \sin\psi \text{ and } \rho_0 = a/s\end{aligned}\right\} \quad . \quad . \quad (52)$$

The profile of the outer conductor thus defined is a tractrix.

So far, the effect of the field penetrating the resistive film into the enclosure formed by it has been ignored. For a uniform coaxial system it is shown in Section 7.1 that the effect is to impose a small negative inductance on the film resistance, which can be compensated by a small increase in the radius of the outer conductor. By assuming that the same correction applies to a tapered mount, it follows from eqn. (87) that the right-hand side of eqn. (51) should be multiplied by the factor $(1 + \epsilon_r s^2/4)$ to compensate for the field penetrating the film, where ϵ_r is the relative permittivity of the enclosure compared with that of the coaxial space.

(3.2) Lead-in Cones

To avoid discontinuity at the input to the resistor mount, a pair of conical conductors are used, the outer cone being a natural continuation of the tapered outer and the inner cone being of semi-angle ψ (Fig. 8). Both cones have a common apex. In a lossless conical line with a common apex the wavefronts are parts of spheres, and a spherical co-ordinate system, $u_1 = \theta$, $u_2 = \phi$ and $u_3 = \rho$ (Fig. 9) conforms with the field. For this, $h_1 = \rho$, $h_2 = \rho\sin\theta$ and $h_3 = 1$, so that h_3 and h_1/h_2 are independent of u_3. Eqn. (36) gives

$$Z_c = \frac{\zeta}{2\pi}\int_{\theta_i}^{\theta_0}\frac{d\theta}{\sin\theta} = \frac{\zeta}{2\pi}\log_\varepsilon\left[\frac{\tan(\theta_0/2)}{\tan(\theta_i/2)}\right] \quad . \quad . \quad (53)$$

in which the semi-angle θ_i of the inner cone is set equal to ψ for the lead-in cones. Comparison with eqn. (49) shows that there will be conformity at the outer conductor when the characteristic resistances are equated.* The angle of the inner cone ensures the correct direction of the lines of electric force at the entry to the resistor surface.

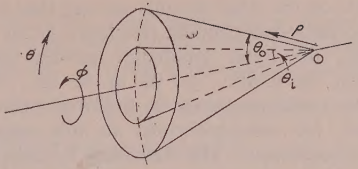


Fig. 9.—Spherical co-ordinate system for a conical coaxial line.

When a conical coaxial line is joined to a cylindrical coaxial line, a spherical wave has to be transformed into a plane wave. The equivalent network of the resulting discontinuity is a low-pass π-section of zero-frequency characteristic impedance equal to the nominal impedance of the line. With small cone angles ($\theta_0 < 15°$) and with diameters within about 1·4 in, this impedance does not change appreciably at any frequency up to 4 000 Mc/s.

(4) EXPERIMENTAL RESISTOR MOUNT

To check the design formulated in the preceding Section, some 24·3 ohm coaxial resistor mounts were constructed, using cracked-carbon-film resistors (Fig. 10). The nominal design was based on eqns. (51) and (52), no correction for the field penetrating the resistive film being made. In the best experimental mount the profile of the outer was within ±0·0005 in of the calculated form, but the diameter of the resistor varied by ±0·002 in. The admittance was measured at frequencies from 5 to 250 Mc/s by a precision admittance bridge and from 560 Mc/s to 3 450 Mc/s by a slotted line. The results are given in the Table. Correction for the field penetration would reduce the small phase angles shown in the Table.

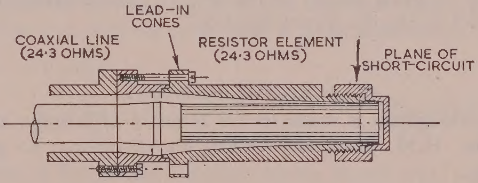


Fig. 10.—Section of experimental "terminal" resistor mount, of 24·3 ohms.
Length of resistor, 2·35 in.

Table 1

G_{dc} = 41·15 Millimhos

Frequency	G/G_{dc}	$B/G(= \tan\phi)$
Mc/s		
5	1·005 ± 0·001	0·001 ± 0·0003
10	1·005 ± 0·001	0·0008 ± 0·0003
30	1·003 ± 0·001	0·0014 ± 0·0003
150	1·009 ± 0·002	0·0036 ± 0·001
200	1·009 ± 0·002	0·0024 ± 0·001
250	1·000 ± 0·002	0·0023 ± 0·001
560	0·99 ± 0·01	0·013 ± 0·01
2 620	0·99 ± 0·01	−0·04 ± 0·05
2 960	1·01 ± 0·01	−0·04 ± 0·05
3 450	0·99 ± 0·01	−0·05 ± 0·05

* It is evident from a comparison between eqns. (53) and (49) that the tractrix profile of the outer conductor will result from a construction according to Fig. 1, using conical line sections in place of cylindrical line sections.

Experiments on the other mounts showed that, to obtain a substantially frequency-independent resistance at all frequencies up to 4 000 Mc/s, the outer conductor profile and the resistor diameter must be within ±0·001 in of the nominal values. Moreover, the resistance must be uniformly distributed within narrow limits, for non-uniformity in this respect introduces appreciable susceptance. The frequency limitation is set by the start of supplementary modes of propagation, given by $\lambda \simeq \pi(a + b)$, where a and b are the maximum inner and outer conductor radii and λ is the free-space wavelength in the coaxial dielectric.

Because of the small field penetration into the cavity formed by the resistor, it might be thought possible to insert thermocouples or similar devices into the cavity to measure power. This is not possible, however, because the insertion of a conductor into the centre of the system completely changes the electric-field configuration and seriously affects the resistance. A better way to measure the power dissipated in a terminating load is to construct a dissipative attenuator from "through" and "terminal" versions of the foregoing design of resistor mount, and use an aperiodically mounted pair of thermistors as a bolometer to terminate the attenuator. By this means u.h.f. power up to 60 watts can be dissipated in a known resistance and measured, using ordinary film resistors in an attenuator of 45 dB, with a thermistor mount which will handle 2 mW of radio-frequency power. Applications of the simple "terminal" resistor mount to the field of measurement are too well known to require special mention.

(5) ACKNOWLEDGMENT

Acknowledgment is made to the Ministry of Supply, and the Controller, H.M. Stationery Office, for permission to publish the paper.

(6) REFERENCES

(1) Crosby, D. R., and Pennypacker, C. H.: "Radio Frequency Resistors as Uniform Transmission Lines," *Proceedings of the Institute of Radio Engineers*, 1946, **34**, p. 62.

(2) Kohn, C. T.: "The Design of a Radio Frequency Coaxial Resistor," *Proceedings I.E.E.*, Monograph No. 83, November, 1953 (**101**, Part IV, p. 146).

(3) Montgomery, C. C., *et al.*: "Technique of Microwave Measurements," M.I.T. Radiation Laboratory Series No. 11, p. 725 (McGraw Hill, New York, 1947).

(4) Heaviside, O.: "Electromagnetic Theory" (Spon, London, 1951), Vol. 1, p. 386.

(5) Bromwich, T. J. I'A.: "Electromagnetic Waves," *Philosophical Magazine*, 1919, **38**, p. 143.

(6) See, for example, Piaggio, H. T. H.: "Differential Equations" (Bell, London, 1933), p. 109.

(7) British Patent Specification No. 670339.

(8) McLachlan, N. W.: "Bessel Functions for Engineers" (University Press, Oxford, 1941).

(7) APPENDIX

(7.1) The Principal Mode in a Uniform Coaxial Line with a Resistive-Film Inner Conductor

A uniform thin resistive film of thickness d in the form of a cylinder of outer radius a is surrounded by a perfectly conducting cylinder of inner radius b. This is described by cylindrical co-ordinates r, ϕ, z, with the z-axis along the axis of the system. The field is divided into three regions: I from $r = b$ to $r = a$; II from $r = a$ to $r = a - d$; and III from $r = a - d$ to $r = 0$. Regions I and III are of loss-free dielectric, while region II is of conductivity σ, which is such that displacement currents in the region are negligible. Field components are restricted to E_r, E_z and H_ϕ, and for regions I and III Maxwell's equations become

$$\left.\begin{aligned} \partial E_r/\partial z - \partial E_z/\partial r &= -j\omega\mu H_\phi \\ -\partial H_\phi/\partial z &= j\omega\epsilon E_r \\ 1/r\partial(rH_\phi)/\partial r &= j\omega\epsilon E_z \end{aligned}\right\} \quad . \quad . \quad . \quad (54)$$

In region II, in which $\sigma \gg \omega\epsilon$, there obtains

$$\left.\begin{aligned} \partial E_r/\partial z - \partial E_z/\partial r &= -j\omega\mu H_\phi \\ -\partial H_\phi/\partial z &= \sigma E_r \\ 1/r.\partial(rH_\phi)/\partial r &= \sigma E_z \end{aligned}\right\} \quad . \quad . \quad . \quad (55)$$

but since $d \ll a$, the last equation may be written $\partial H_\phi/\partial r \simeq \sigma E_z$ with sufficient accuracy.

Eqns. (54) may be solved by Bromwich's method,[5] in which the solution is the sum of two partial solutions, one being derived when $E_z = 0$ and the other when $H_z = 0$. Here, H_z is identically zero, and only the solution with $H_z = 0$ need be considered. The solution is obtained through a potential function U which satisfies the equation

$$1/r.\partial(r\partial U/\partial r)/\partial r + \partial^2 U/\partial z^2 + \omega^2\epsilon\mu U = 0 \quad . \quad (56)$$

The field components follow from the relationships

$$\left.\begin{aligned} E_r &= \partial^2 U/\partial z\partial r \\ E_z &= \partial^2 U/\partial z^2 + \omega^2\epsilon\mu U \\ H_\phi &= -j\omega\epsilon\partial U/\partial r \end{aligned}\right\} \quad . \quad . \quad . \quad (57)$$

Eqn. (56) may be solved by separating the variables, writing $U = \mathscr{R}(r)Z(z)$, resulting in two equations

$$d^2\mathscr{R}/dr^2 + 1/r.d\mathscr{R}/dr - n^2\mathscr{R} = 0 \quad . \quad . \quad . \quad (58)$$

$$d^2Z/dz^2 + (k^2 + n^2)Z = 0 \quad . \quad . \quad . \quad . \quad (59)$$

in which $k = \omega(\epsilon\mu)^{1/2}$ and n is a constant. The solution of eqn. (58) is

$$\mathscr{R} = C_1 I_0(nr) + C_2 K_0(nr) \quad . \quad . \quad . \quad . \quad (60)$$

in which I_0 and K_0 are modified Bessel functions of zero order and of the first and second kind, respectively. Let

$$-\gamma^2 = k^2 + n^2 \quad . \quad . \quad . \quad . \quad (61)$$

The solution of eqn. (59) is $z = C_3\varepsilon^{\gamma z} + C_4\varepsilon^{-\gamma z}$, from which it appears that γ is the propagation coefficient of a wave travelling in the z-direction. If propagation be restricted to the positive z direction, then $C_3 = 0$. On multiplying z, thus restricted, by $\mathscr{R}$ and absorbing C_4 in C_1 and C_2 there results

$$U = [C_1 I_0(nr) + C_2 K_0(nr)]\varepsilon^{-\gamma z} \quad . \quad . \quad . \quad (62)$$

With eqn. (62) in eqns. (57) there results

$$H_\phi = [C_1 I_1(nr) - C_2 K_1(nr)]\varepsilon^{-\gamma z} \quad . \quad . \quad . \quad . \quad (63)$$

$$E_r = (\gamma/j\omega\epsilon)[C_1 I_1(nr) - C_2 K_1(nr)]\varepsilon^{-\gamma z} \quad . \quad . \quad (64)$$

$$E_z = (n/j\omega\epsilon)[C_1 I_0(nr) + C_2 K_0(nr)]\varepsilon^{-\gamma z} \quad . \quad . \quad (65)$$

in which $I_1(x) = dI_0(x)/dx$ and $K_1(x) = -dK_0(x)/dx$ and certain constants have been changed. Eqns. (63)–(65) apply to regions I and III. For region II the last two (simplified) of eqns. (55) are substituted in the first, giving

$$\partial^2 H_\phi/\partial r^2 + \partial^2 H_\phi/\partial z^2 + k_2^2 H_\phi = 0$$

in which $k_2^2 = -j\omega\mu\sigma$. The solution for forward propagation only is found, by separating the variables, to be

$$H_\phi = (C_3\varepsilon^{n_2 r} + C_4\varepsilon^{-n_2 r})\varepsilon^{-\gamma_2 z} \quad . \quad . \quad . \quad (66)$$

with $-\gamma_2^2 = k_2^2 + n_2^2$, n_2 being a constant. From eqn. (66) and the simplified form of eqns. (55) there follow

$$E_r = (\gamma_2/\sigma)(C_3\varepsilon^{n_2 r} + C_4\varepsilon^{-n_2 r})\varepsilon^{-\gamma_2 z} \quad . \quad . \quad (67)$$

$$E_z = (n_2/\sigma)(C_3\varepsilon^{n_2 r} - C_4\varepsilon^{-n_2 r})\varepsilon^{-\gamma_2 z} \quad . \quad . \quad (68)$$

In eqns. (66) and (68) it is expedient to write $r - a$ for r and absorb $\varepsilon^{\pm n_2 a}$ into the constants.

The boundary condition at $r = b$ is simply that the electric vector is normal to the surface, i.e. $E_z = 0$. From eqn. (65) follows

$$C_1 = -C_2\mathrm{K}_0(n_1 b)/\mathrm{I}_0(n_1 b)$$

in which the suffix 1 on n refers to region I. Then the field strengths in region I are

$$H_\phi = -n_1 C_2[\mathrm{K}_0(n_1 b)\mathrm{I}_1(n_1 r)/\mathrm{I}_0(n_1 b) + \mathrm{K}_1(n_1 r)]\varepsilon^{-\gamma_1 z} \quad . \quad (69)$$

$$E_r = -(\gamma_1 n_1 C_2/j\omega\epsilon_1)[\mathrm{K}_0(n_1 b)\mathrm{I}_1(n_1 r)/\mathrm{I}_0(n_1 b) + \mathrm{K}_1(n_1 r)]\varepsilon^{-\gamma_1 z} \quad . \quad . \quad . \quad . \quad (70)$$

$$E_z = -(n_1^2 C_2/j\omega\epsilon_1)[\mathrm{K}_0(n_1 b)\mathrm{I}_0(n_1 r)/\mathrm{I}_0(n_1 b) - \mathrm{K}_0(n_1 r)]\varepsilon^{-\gamma_1 z} \quad . \quad . \quad . \quad . \quad (71)$$

In region II, eqns. (66), (67) and (68) apply. In region III, C_2 must vanish because K_0 and K_1 are infinite at $r = 0$. Then

$$H_\phi = C_5\mathrm{I}_1(n_3 r)\varepsilon^{-\gamma_3 z} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (72)$$

$$E_r = (\gamma_3 C_5/j\omega\epsilon_3)\mathrm{I}_1(n_3 r)\varepsilon^{-\gamma_3 z} \quad . \quad . \quad . \quad (73)$$

$$E_z = (n_3 C_5/j\omega\epsilon_3)\mathrm{I}_0(n_3 r)\varepsilon^{-\gamma_3 z} \quad . \quad . \quad . \quad (74)$$

in which the suffix 3 denotes the values for region III. The boundary conditions at $r = a$ and $a - d$ are that the tangential components of $\boldsymbol{E}$ and $\boldsymbol{H}$ immediately on either side of the boundary are equal. Then, from eqns. (66)–(71),

$$\gamma_1 = \gamma_2 = \gamma_3 \, (= \gamma)$$

$$n_1[\mathrm{K}_0(n_1 b)\mathrm{I}_1(n_1 a)/\mathrm{I}_0(n_1 b) + \mathrm{K}_1(n_1 a)]C_2 + C_3 + C_4 = 0 \quad . \quad . \quad . \quad . \quad . \quad (75)$$

$$n_1^2/j\omega\epsilon_1)[\mathrm{K}_0(n_1 b)\mathrm{I}_0(n_1 a)/\mathrm{I}_0(n_1 b) - \mathrm{K}_0(n_1 a)]C_2 + (n_2/\sigma)C_3 - (n_2/\sigma)C_4 = 0 \quad . \quad (76)$$

$$\varepsilon^{-n_2 d}C_3 + \varepsilon^{n_2 d}C_4 - \mathrm{I}_1(n_3 a')C_5 = 0 \quad . \quad . \quad . \quad (77)$$

$$(n_2/\sigma)\varepsilon^{-n_2 d}C_3 - (n_2/\sigma)\varepsilon^{n_2 d}C_4 - [n_3\mathrm{I}_0(n_3 a')/j\omega\epsilon_3]C_5 = 0 \quad (78)$$

n which $a' = a - d$. Eqns. (75)–(78) from a set of simultaneous equations for three of the four C's, the remaining C being arbitrary; n_2 and n_3 are given in terms of n_1 by

$$n_2^2 = n_1^2 + k_1^2 + j\omega\mu\sigma \quad . \quad . \quad . \quad . \quad . \quad . \quad (79)$$

$$n_3^2 = n_1^2 - k_1^2[(\mu_3\epsilon_3/\mu_1\epsilon_1) - 1] \quad . \quad . \quad . \quad (80)$$

To ensure compatibility of eqns. (75)–(78), the determinant of he array of coefficients of the four C's must be zero. The determinant is expanded, setting $e^{\pm n_2 d} \simeq 1 \pm n_2 d$, because $n_2 d \ll 1$, and setting $a' \simeq a$. The following approximations to he Bessel functions,[8] valid for arguments small compared with unity, are then substituted in the result:

$$\left.\begin{aligned} \mathrm{K}_0(x) &= (-\log_\varepsilon x + 0{\cdot}116)(1 + x^2/4) + x^2/4 \\ \mathrm{K}_1(x) &= 1/x + (x/2)(\log_\varepsilon x - 0{\cdot}116) - x^2/4 \\ \mathrm{I}_0(x) &= 1 + x^2/4; \; I_1(x) = x/2 \end{aligned}\right\} \quad . \quad (81)$$

The result is solved for n_1, after eliminating n_2 and n_3 by eqns. (79) and (80) and making some approximations appropriate to the practical application. The first-order approximation is

$$n_1^2 = -j\omega\epsilon_1/[\sigma da \log_\varepsilon (b/a)] \quad . \quad . \quad . \quad . \quad (82)$$

and the second-order approximation* is

$$n_1^2 = \frac{-j\omega\epsilon_1}{\sigma da \log_\varepsilon (b/a)} - \left[\frac{\omega\epsilon_1}{\sigma d \log_\varepsilon (b/a)}\right]^2 \left[\frac{\epsilon_3 \log_\varepsilon (b/a)}{2\epsilon_1} + \frac{b^2/a^2 - 1}{4 \log_\varepsilon (b/a)} - \frac{1 + \log_\varepsilon (b/a)}{2}\right] \quad . \quad (83)$$

In the simple transmission-line theory of the system

$$\gamma^2 = -\omega^2 L_0 C_0 + j\omega C_0 R_0 \quad . \quad . \quad . \quad . \quad (84)$$

in which

$$L_0 = (\mu_1/2\pi)\log_\varepsilon (b/a)$$
$$C_0 = 2\pi\epsilon_1/\log_\varepsilon (b/a)$$
$$R_0 = 1/(2\pi\sigma ad)$$

In the field theory eqn. (61) gives $\gamma^2 = -k_1^2 - n_1^2$, and with eqn. (83) and the above values of L_0, C_0 and R_0, there results

$$\gamma^2 = -\omega^2 L_0 C_0 + j\omega C_0 R_0 + (\omega C_0 R_0 a)^2[\epsilon_3 \log_\varepsilon (b/a)/2\epsilon_1 + \ldots] \quad . \quad . \quad . \quad . \quad . \quad (85)$$

On comparing this with eqn. (84) it is seen that the second-order correction is equivalent to an increment in inductance

$$\Delta_1 L_0 = -C_0 R_0^2 a^2 \left[\frac{\epsilon_3 \log_\varepsilon (b/a)}{2\epsilon_1} + \frac{b^2/a^2 - 1}{4 \log_\varepsilon (b/a)} - \frac{1 + \log_\varepsilon (b/a)}{2}\right] \quad . \quad (86)$$

The term in the brackets containing the factor ϵ_3/ϵ_1 must arise from the field which penetrates into region III, for if $\epsilon_3 = 0$ the enclosure would support no flux. The other two terms in the brackets must result from the axial component in region I. Therefore the effect of penetration into region III is a small increment of inductance

$$\Delta L_0 = -\pi\epsilon_3(R_0 a)^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (87)$$

which results from (86) and the value of C_0 given above.

From eqns. (69)–(74), with appropriate approximations, the real parts of the expressions for E_r and E_z may be obtained, and from them the lines of electric force may be plotted. The general form of the field thus obtained is sketched in Fig. 11.

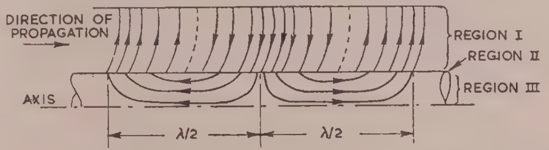


Fig. 11.—Electric flux in a coaxial line with a resistive-film inner conductor.

(7.2) The Orthogonal Curvilinear Co-ordinate System Relating to a Resistor Mount

Refer to Fig. 12. The cross-sections of the surfaces $u_3 =$ a constant are stipulated to be parts of circles of equal radii ρ_0 with centres on the axis. Let the numerical value of u_3 be given by the distance w_1 along the axis, measured from the point where the part circle $u_3 = 0$ cuts the axis. Then

$$r = +[\rho^2 - (\rho_0 + w_1 - u_3)^2]^{1/2} \quad . \quad . \quad . \quad (88)$$

* It is assumed that $\mu_1 = \mu_2 = \mu_3$ in these results.

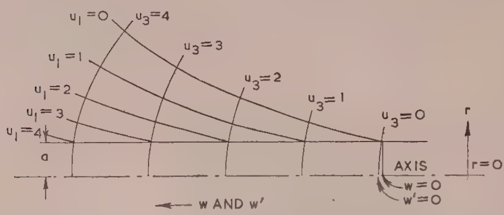


Fig. 12.—Orthogonal co-ordinate system for a tapered resistor mount.

Differentiation and elimination of u_3 gives

$$(dr/dw_1)_{u_3=\text{constant}} = -(\rho_0^2 - r^2)^{1/2}/r \quad . \quad . \quad (89)$$

Orthogonal trajectories to the family of part-circles (88) are given by the relationship

$$(dr/dw_1)_{u_1=\text{constant}} = -1/(dr/dw_1)_{u_3=\text{constant}}$$

which with eqn. (89) leads to the integral

$$w_1 = \int \frac{(\rho_0^2 - r^2)^{1/2}}{r} dr = (\rho_0^2 - r^2)^{1/2} - \rho_0 \log_\varepsilon \left[\frac{\rho_0 + (\rho_0^2 - r^2)^{1/2}}{r}\right] + C \quad . \quad . \quad . \quad . \quad (90)$$

This describes a family of tractrices, the constant C distinguishing the individual members. Let u_3 also be numerically equal to the distance w measured along the resistor surface (see Fig. 12). Then eqn. (88) with $r = a$ gives

$$w_1 = -\rho_0 + (\rho_0^2 - a^2)^{1/2} + w \quad . \quad . \quad . \quad (91)$$

When $r = a$, $w = u_3 = u_1$ and putting u_1 for w in eqn. (91) and substituting for w_1 in eqn. (90) with $r = a$, gives

$$C = u_1 - \rho_0 + \rho_0 \log_\varepsilon \{[\rho_0 + (\rho_0^2 - a^2)^{1/2}]/a\} \quad . \quad (92)$$

Again, with eqns. (91) and (92) in eqn. (90) there is obtained

$$u_1 = w + (\rho_0^2 - a^2)^{1/2} - (\rho_0^2 - r^2)^{1/2} - \rho_0 \log_\varepsilon \{[\rho_0 + (\rho_0^2 - a^2)^{1/2}]/a\} + \rho_0 \log_\varepsilon \{[\rho_0 + (\rho_0^2 - r^2)^{1/2}]/r\} \quad . \quad (93)$$

On substituting from eqn. (91) for w_1 in eqn. (88), there results

$$u_3 = w + (\rho_0^2 - a^2)^{1/2} - (\rho_0^2 - r^2)^{1/2} \quad . \quad . \quad (94)$$

By analogy with the relationship $\sin \psi = a/\rho_0$ (Section 3.1) let $\sin \theta = r/\rho_0$ define another angle θ. Define two quantities ξ and ξ_i by $\tan(\theta/2)$ and $\tan(\psi/2)$ respectively, and note the identities $\tan(\theta/2) = \sin \theta/(1 + \cos \theta) = (1 - \cos \theta)/\sin \theta$; $\cos \theta = (1 - \xi^2)/(1 + \xi^2)$ and $\sin \theta = 2\xi/(1 + \xi^2)$. On substituting ξ and ξ_i for r and a in eqns. (93) and (94) there is obtained

$$u_1 = w + \rho_0[(1 - \xi_i^2)/(1 + \xi_i^2) - (1 - \xi^2)/(1 + \xi^2) - \log_\varepsilon (\xi/\xi_i)] \quad . \quad (95)$$

$$u_3 = w + \rho_0[(1 - \xi_i^2)/(1 + \xi_i^2) - (1 - \xi^2/)(1 + \xi^2)] \quad (96)$$

By subtraction $$u_3 - u_1 = \rho_0 \log_\varepsilon (\xi/\xi_i) \quad . \quad . \quad . \quad . \quad (97)$$

or $$\xi = \xi_i \exp [(u_3 - u_1)/\rho_0] \quad . \quad . \quad . \quad . \quad (98)$$

To obtain the relationship between r and w it is noted from the definition of θ that $r = \rho_0 \sin \theta$, giving $r = 2\rho_0\xi/(1 + \xi^2)$, in which ξ is given by eqn. (98). At the outer conductor surface $u_1 = 0$, and when following a curve $u_3 =$ a constant, then $u_3 = w$ numerically. Since $\xi_i = [1 - (1 - s^2)^{1/2}]/s \simeq (s/2)(1 + s^2/4)$, where $s = a/\rho_0$, there is obtained the approximate relation

$$r \simeq \frac{a(1 + s^2/4) \exp (w/\rho_0)}{1 + (s^2/4) \exp (2w/\rho_0)} \quad . \quad . \quad . \quad . \quad (99)$$

This gives r at the axial distance corresponding to the point where the curve $u_3 = w$ cuts the outer conductor. This point is displaced axially Δw from the point w on the resistor surface, where

$$\Delta w = \rho_0(\cos \psi - \cos \theta) \simeq (as/2)[\exp (2w/\rho_0) - 1] \quad . \quad (100)$$

Eqns. (99) and (100) determine the profile of the outer conductor sufficiently accurately for most purposes. Alternatively, w is given exactly as a function of r by eqn. (93) with $u_1 = 0$. This is the equation to a tractrix.

To obtain h_1, h_2 and h_3, note that u_2 may be expressed in terms of the two co-ordinates x and y in a plane perpendicular to z or w where $x = r \cos u_2$ and $y = r \sin u_2$. Then

$$x = (2\rho_0\xi \cos u_2)/(1 + \xi^2) \quad \text{and} \quad y = (2\rho_0\xi \sin u_2)/(1 + \xi^2)$$

which together with eqn. (96) enable h_1, h_2 and h_3 to be calculated from eqn. (5) and two like it. The results are

$$\left.\begin{aligned} h_1 &= 2\xi/(1 + \xi^2) \\ h_2 &= 2\xi\rho_0/(1 + \xi^2) \\ h_3 &= (1 - \xi^2)/(1 + \xi^2) \end{aligned}\right\} \quad . \quad . \quad . \quad (101)$$

where ξ is given by eqn. (98). Then $h_1/h_2 = 1/\rho_0$ is constant and h_3 varies very slowly with u_3, provided that $(u_3 - u_1)/\rho_0$ is not large and $a/\rho_0 \ll 1$. Thus the conditions for a valid general principal wave are approximately satisfied.

621.313.333: 621.3.016

Monograph No. 133 U
May 1955

SECOND-ORDER TORQUE COMPONENTS IN THE SCHRAGE MOTOR OPERATING AT SYNCHRONOUS SPEED

By I. THOMAS, B.Sc., Student.

(*The paper was first received* 16*th August*, 1954, *and in revised form* 7*th February*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in May*, 1955.)

SUMMARY

The operation of a Schrage motor at synchronous speed is considered, and on the assumption that the air-gap flux at this speed contains space harmonics, an expression for the torque in terms of flux-axis position is derived. This shows that the torque consists of the main component, which is constant for constant brush separation, and variable components depending on (*a*) unbalance in the secondary circuits owing to errors in brush-separation angles and/or unequal secondary resistances, and (*b*) the assumed space harmonics of flux. The variable torque components give rise to superimposed speed oscillations and primary-current hunting, when the average speed is just above or below synchronism.

An experimental method of investigating the torque is then given, together with two methods of measuring the harmonic content of the air-gap flux. The latter measurements are used to predict the amount by which the total torque varies with the air-gap flux-axis position at synchronous speed.

Possible errors are discussed, and an approximate correction for non-linearity due to the brush-contact effect is derived and applied.

Experimental results given for a 2-phase machine show reasonable agreement with the theory.

LIST OF SYMBOLS

Unless otherwise stated, the following symbols apply when the Schrage motor is operating at synchronous speed:

A_0 = Factor showing the effect of flux harmonics on constant torque component developed.

A_a = Factor showing extent of *a*th harmonic of torque.

a = Order of torque harmonic.

B = Resultant air-gap flux density, Wb/m^2.

B_n = Maximum air-gap flux density of *n*th flux harmonic, Wb/m^2.

d = Difference between brush-spread/secondary-resistance ratios for 2-phase machine (using R_1 and R_2).

d' = Difference between brush-spread/secondary-resistance ratios for 2-phase machine (using R_1' and R_2').

f_0 = Sum of brush-spread/secondary-resistance ratios for *u*-phase machine.

f_a, g_a = Functions of brush-spread/secondary-resistance ratios, depending on number of secondary phases.

i = Current flowing in secondary circuit.

i_u = Current flowing in circuit of *u*th secondary phase.

i_{um} = Current flowing in circuit of *u*th secondary phase owing to voltage produced by rotation in *m*th harmonic flux.

$K_n = 2K_v k_n \sin \frac{1}{2}\beta_1$.

$2T_t(= k)$ = Maximum torque produced when a current of 1 amp flowing in one secondary circuit reacts with the fundamental air-gap flux.

$K_v = B_1 \omega_s NYl \cos \gamma / \pi \sin \frac{1}{2}\psi$.

$k_n, \bar{k}_n, k_m$ = Ratio of *n*th (*m*th) harmonic to fundamental brush voltage (function of brush spread).

k_n'' = Ratio of maximum *n*th harmonic to maximum fundamental brush voltage (independent of brush shift).

l = Length of stator-coil sides, m.

N, N' = Number of turns per coil of commutator and stator windings, respectively.

n, m = Order of air-gap flux space harmonics.

r = Radial distance of stator-coil sides, m.

R = Total resistance of secondary circuit.

R_u = Total resistance of circuit of *u*th secondary phase.

R_u' = Ratio of maximum fundamental voltage to maximum fundamental current for *u*th secondary phase.

s = Sum of brush-spread/secondary-resistance ratios for 2-phase machine (using R_1 and R_2).

s' = Sum of brush-spread/secondary-resistance ratios for 2-phase machine (using R_1' and R_2').

T = Total torque developed, newton-metres.

T_0 = Constant component of total torque.

T_a = *a*th harmonic components of total torque.

T_u = Total torque on stator winding S_u.

T_{umn} = Torque on stator winding S_u when carrying curren i_{um} in *n*th harmonic flux.

U = Total number of secondary phases.

u = Numeral applied to particular secondary phase.

V = Brush voltage.

V_{max} = Maximum brush voltage.

V_u = Brush voltage of *u*th secondary phase.

V_{um} = Brush voltage of *u*th secondary phase due to rotation in *m*th harmonic flux.

V_n' = *n*th time harmonic brush voltage at standstill.

Yl = Pole area, m^2.

$\alpha(= \alpha_1)$ = Angle between axes of air-gap flux and secondary circuit 1, rad.

α_u = Angle between axis of air-gap flux and correct axis of stator winding of *u*th secondary phase, rad.

β = Brush spread, rad.

$\beta_u, \bar{\beta}_u$ = Brush spread for *u*th secondary phase, rad.

β' = Stator-winding spread, rad.

$2\gamma, 2\gamma'$ = Angular short-chording of commutator and stator windings, respectively, rad.

δ = Brush shift, rad.

η_a = arc tan (g_a/f_a), rad.

θ = Position round air-gap.

θ' = Position round air-gap, standstill conditions.

σ_n, σ_n' = Ratio of *n*th harmonic winding factor to fundamental winding factor for commutator and stator windings, respectively.

$\bar{\sigma}_n = (\sigma_n'/\sigma_n)$.

Φ = Total flux per pole, Wb.

ψ, ψ' = Slot pitch of commutator and stator windings, respectively.

ω = Synchronous rotor speed, rad/sec.

ω_s = Rotor speed, rad/sec.

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Thomas is with The General Electric Co., Ltd., and was formerly at King's College, University of Durham.

(1) INTRODUCTION

It is well known that the Schrage motor is capable of showing periodic variations in speed and line currents when its average speed is near synchronism. Such hunting is at least partly explained by the possibility of the output voltage from the commutator being asymmetrical, or connected to an unbalanced system of impedances. Asymmetrical commutator voltages can be caused by errors in brush positioning; this effect is mentioned in a paper by Arnold,[1] although not thoroughly dealt with. The effect of unbalanced secondary circuits in general is dealt with qualitatively by Adkins and Gibbs.[2]

Speed and current hunting caused by either or both of the above effects possesses a frequency which is twice the slip frequency. In the course of investigation, however, it became clear that, for the particular machine under test, hunting of a frequency four times that of the slip is also possible. Since the Schrage motor is an asynchronous machine, any speed variation is not of the phase-swinging type associated with synchronous machines, but simply variation about an average speed. The speed variations are accompanied by corresponding periodic variations in the torque developed by the machine, and the present paper sets out to investigate the torque characteristics; this is best done by considering the machine operating at synchronous speed. It is seen that the torque characteristics which can produce speed hunting depend not only on the factors mentioned above, but also on space harmonics in the air-gap flux distribution of the machine.

(2) OPERATION OF THE SCHRAGE MOTOR

(2.1) General

The Schrage motor is an inverted induction motor and a frequency changer combined in the one machine. A rotating flux is produced by polyphase currents flowing in the primary winding on the rotor. With the rotor in motion, this flux induces voltages of slip frequency in secondary windings on the stator, and voltages of supply frequency in the conductors of a commutator winding housed in the rotor slots. Owing to the effect of the commutator the brush voltages are always of slip frequency, so that they can be injected (in correct phase sequence) into the secondary circuits. Speed and/or power-factor control is then obtained by suitable movement of the brushes on the commutator.

(2.1.1) Operation at Synchronous Speed.

Since the primary winding is on the rotor, the speed of rotation of the air-gap flux relative to the stator is the difference between synchronous speed and rotor speed. Thus when the rotor is running at synchronous speed, the air-gap flux becomes stationary. No voltage will then be generated in the stator windings, while the generation of voltage at the commutator can be compared with that of a d.c. machine. The position of the air-gap flux, although stationary, does, however, depend on the instantaneous position of the rotor measured in electrical degrees at time intervals equal to the periodic time of the mains supply. This position will be referred to as the "synchronous position" of the rotor.

(2.2) Torque Production at Synchronous Speed

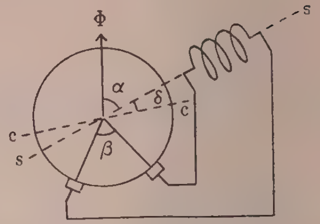


Fig. 1.—Operation with one secondary circuit connected.
cc. = Commutator winding axis.
ss. = Stator winding axis.

Some idea of the torque condition of a Schrage motor operating at synchronous speed can be obtained by considering such a machine with one secondary circuit only in operation. By "secondary circuit" is meant the usual combination of a stator winding and its corresponding commutator circuit, as shown for the 2-pole case in Fig. 1. Since the speed is synchronous the air-gap flux Φ, due to polyphase primary currents is stationary. If it is assumed that Φ is constant with respect t time and position and sinusoidally distributed in the air-gap the open-circuit voltage at the brushes is

$$V = V_{max} \sin \tfrac{1}{2}\beta \sin (\alpha + \delta)$$

where β is the brush separation, α is the angle between stato winding and flux axes, and δ is the angle between the magneti axes of the stator winding and the commutator winding, and i known as "brush shift." If R is the total resistance of th secondary circuit, then, assuming that the reaction of th secondary circuit is negligible, the secondary current is

$$i = (V_{max}/R) \sin \tfrac{1}{2}\beta \sin (\alpha + \delta)$$

The torque exerted on the stator winding carrying this currer is proportional to the product of the current and the componer of Φ in quadrature with the winding axis. Thus the torque is

$$T = (kV_{max}/R) \sin \tfrac{1}{2}\beta \sin (\alpha + \delta) \sin \alpha$$
$$= \tfrac{1}{2}(kV_{max}/R) \sin \tfrac{1}{2}\beta [\cos \delta - \cos (2\alpha + \delta$$

It causes motion by reaction on the rotor. The expressio shows that the machine considered will operate synchronousl i.e. for a change in mechanical load it is merely necessary for th axis of Φ to find a new position in the air-gap, provided that th torque required is not greater than

$$\tfrac{1}{2}(kV_{max}/R) \sin \tfrac{1}{2}\beta [\cos \delta + 1]$$

With all polyphase secondary circuits in operation, and t system symmetrical and balanced, the combination of t cos $(2\alpha + \delta)$ terms will have no resultant. Thus with α, β and all constant, the total torque developed at synchronous spe will be single-valued, and the machine will run constantly this speed only, provided that the mechanical load is constant the appropriate value.

However, with any form of unbalance in the seconda circuits, the cos $(2\alpha + \delta)$ terms when combined will have finite resultant. With δ, β and R all constant, the torque w then consist of a constant component and a component dependi on the position taken by the flux axis in the air-gap. Such machine is capable of operating at synchronous speed over limited torque range, the limits being determined by the exte of unbalance in the secondary circuits.

(2.3) Operation at Speeds nearing Synchronism

When the torque required by a machine with unbalanc secondary circuits is just outside the range for which synchrono running is possible, synchronism is lost, and the main fl rotates slowly in the air-gap. The component of torque th oscillates, which tends to produce periodic variations in spe. From the expressions given in the preceding Section, it is cl that the frequency of the periodic speed variations, if they ex is twice the slip frequency. The purpose of the paper is to investigate the nature of the torques which produce these speed variations, rather than the nature of the variations themselves.

(3) AN EXPRESSION FOR TORQUE AT SYNCHRONISM

There are second-order torque components in ordinary induction motors known to be caused by time harmonics in the supply and space harmonics due to winding distribution[3,4] and the presence of slot openings.[5,6] In order to take harmonics into account when considering the Schrage motor operating at synchronous speed, the component of air-gap flux which is stationary at this speed is assumed not to be sinusoidally distributed, but to contain all odd space harmonics. Harmonic components of the total flux which rotate when the machine is operating at its synchronous speed need not be considered. Thus, that part of the air-gap flux which is stationary is assumed to be

$$B = (B_1 \cos\theta - B_3 \cos 3\theta + B_5 \cos 5\theta \ldots)$$

$$= \sum_{n=1}^{\infty} (-1)^{(n-1)/2} B_n \cos n\theta \quad . \quad (1)$$

where n is a positive odd integer, θ is in electrical radians, and B_n is the maximum flux density of the nth harmonic measured in webers per square-metre.

The B_n's are assumed to be independent of α, the flux-axis position in space, and the whole air-gap flux is assumed to be due to an m.m.f. which is the resultant of the primary-, stator- and commutator-winding m.m.f. These points are discussed further in Section 12.

Finally, for each secondary circuit, the commutator- and stator-winding axes are assumed to be coincident. Errors in brush shift and in stator-winding positions are excluded from the following analysis in order to avoid too complicated a treatment. Brush-spread errors are the most effective in the production of secondary asymmetry, and are therefore dealt with in particular.

With these assumptions, let us consider the uth phase of a machine having a U-phase secondary system. In Fig. 2 $\mathrm{S_u}$ is the stator winding of secondary phase u, and b_u, b_u are the brushes of that phase; their common axis is shown by the broken line. With the axis of the air-gap flux at an angle of α radians to the axis of the secondary circuit 1, the corresponding angle for the secondary circuit u will be $[\alpha + 2\pi(u-1)/U]$ radians.

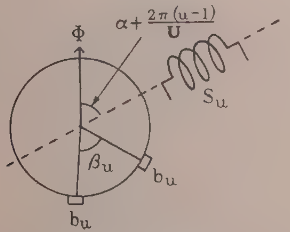


Fig. 2.—Components of the uth secondary phase with zero brush shift.

When the brushes have a separation of β_u radians, the voltage appearing at them due to rotation in the mth harmonic flux $(-1)^{(m-1)/2} B_m \cos m\theta$ is

$$V_{um} = \sin\tfrac{1}{2}\beta_u (2\omega_s NYl \cos\gamma/\pi \sin\tfrac{1}{2}\psi)\sigma_m B_m \sin m[\alpha + 2\pi(u-1)/U]$$

$$= 2K_v k_m \sin\tfrac{1}{2}\beta_u \sin m[\alpha + 2\pi(u-1)/U] \quad . \quad . \quad . \quad . \quad (2)$$

where $\quad K_v = B_1\omega_s NYl \cos\gamma/\pi \sin\tfrac{1}{2}\psi; \; k_m = \sigma_m B_m/B_1$

$$\sigma_m = \frac{\sin\tfrac{1}{2}m\beta_u \sin\tfrac{1}{2}\psi \cos m\gamma}{\sin\tfrac{1}{2}\beta_u \sin\tfrac{1}{2}m\psi \cos\gamma}$$

The coefficient σ_m is, in fact, the ratio of the mth harmonic winding factor to the fundamental winding factor, and the constants are introduced in this particular form so that k_m is given the practical significance of being the ratio of the mth harmonic brush voltage to the fundamental brush voltage.

The current in the secondary circuit u is

$$i_{um} = K_v(\beta_u/R_u)k_m \sin m[\alpha + 2\pi(u-1)/U] \quad . \quad . \quad (3)$$

where R_u is the total resistance round the circuit considered, and $\sin\tfrac{1}{2}\beta_u$ has been replaced by $\tfrac{1}{2}\beta_u$. This is justifiable because β_u is always small when the Schrage motor is working normally at its synchronous speed.

The torque exerted on the stator winding $\mathrm{S_u}$ carrying current i_{um} in the nth harmonic flux density is

$$T_{umn} = [2i_{um}N'rl \sin\tfrac{1}{2}\beta' \cos\gamma'/\sin\tfrac{1}{2}\psi']B_n\sigma'_n \sin n[\alpha + 2\pi(u-1)/U]$$

$$= 2i_{um}T_t\bar{\sigma}_n k_n \sin n[\alpha + 2\pi(u-1)/U] \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where $\quad T_t = B_1 N'rl \sin\tfrac{1}{2}\beta' \cos\gamma'/\sin\tfrac{1}{2}\psi'; \;\; \bar{\sigma}_n = \sigma'_n/\sigma_n;$

Therefore the total torque on the stator winding $\mathrm{S_u}$ is obtained from eqn. (3) by substitution for i_{um} in eqn. (4) and making the appropriate summations.

i.e. $$T_n = K_vT_t(\beta_u/R_u) \sum_{n=1}^{\propto} \sum_{m=1}^{\propto} \bar{\sigma}_n k_n k_m \{\cos(n-m)[\alpha + 2\pi(u-1)/U]$$

$$- \cos(n+m)[\alpha + 2\pi(u-1)/U]\} \quad . \quad (5)$$

The total torque on all the stator coils, which causes rotation by reaction, is thus

$$T = K_vT_t \sum_{u=1}^{U} \Big[(\beta_u/R_u) \sum_{n=1}^{\propto} \sum_{m=1}^{\propto} \bar{\sigma}_n k_n k_m$$

$$\{\cos(n-m)[\alpha + 2\pi(u-1)/U]$$

$$- \cos(n+m)[\alpha + 2\pi(u-1)/U]\}\Big] \quad . \quad (6)$$

This expression shows the presence of torque components depending on α. These will be termed the "space harmonics of torque." Since n and m are always odd, the coefficients of α, which indicate the order of the torque harmonics, will always be even. The fundamental is therefore absent, but there is a zero-order harmonic of torque, which is the torque component independent of α and is, in fact, constant for each particular brush setting. This constant component is obtained by considering the $\cos(n-m)[\alpha + 2\pi(u-1)/U]$ term in eqn. (6) when $n = m$.

This gives $$T_0 = K_vT_t\Big[\sum_{u=1}^{U}(\beta_u/R_u)\Big]\sum_{n=1}^{\propto}\bar{\sigma}_n k_n^2 = K_vT_tf_0A_0 \quad . \quad (7)$$

where $$f_0 = \sum_{u=1}^{U}(\beta_u/R_u) \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

and $$A_0 = \sum_{n=1}^{\propto}\bar{\sigma}_n k_n^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

To obtain an expression for the ath order torque harmonic, the conditions

$$(n-m) = a; \; (m-n) = a; \; (n+m) = a$$

are imposed in turn on eqn. (6), and the relevant terms are selected. This gives

$$T_a = K_v T_t \sum_{u=1}^{U} (\beta_u/R_u)$$
$$\left[\sum_{n=(a+1)}^{\propto} \bar{\sigma}_n k_n k_{(n-a)} + \sum_{n=1}^{\propto} \bar{\sigma}_n k_n k_{(n+a)} - \sum_{n=1}^{(a-1)} \bar{\sigma}_n k_n k_{(a-n)}\right]$$
$$\cos a[\alpha + 2\pi(u-1)/U]$$

or,
$$T_a = K_v T_t [f_a A_a \cos a\alpha - g_a A_a \sin a\alpha] \quad . \quad . \quad (10)$$

where
$$A_a = \left[\sum_{n=(a+1)}^{\propto} \bar{\sigma}_n k_n k_{(n-a)} + \sum_{n=1}^{\propto} \bar{\sigma}_n k_n k_{(n+a)} - \sum_{n=1}^{(a-1)} \bar{\sigma}_n k_n k_{(a-n)}\right] . (11)$$

$$f_a = \sum_{u=1}^{U} (\beta_u/R_u) \cos [2a\pi(u-1)/U] \quad . \quad . \quad (12)$$

and
$$g_a = \sum_{u=1}^{U} (\beta_u/R_u) \sin [2a\pi(u-1)/U] \quad . \quad . \quad (13)$$

To obtain the total torque at synchronous speed, the torque components for all even positive values of a are added and the steady torque given by eqn. (7) is added. This gives

$$T = K_v T_t [(f_0 A_0 + f_2 A_2 \cos 2\alpha + f_4 A_4 \cos 4\alpha \ldots)$$
$$- (g_2 A_2 \sin 2\alpha + g_4 A_4 \sin 4\alpha \ldots)] \quad . \quad (14)$$

The coefficients f_a and g_a are functions of brush spread and secondary resistances, and by using eqns. (12) and (13) they can be readily evaluated. Both the f_a's and g_a's do, in fact, themselves form a periodic series; for all values of a it can be shown that $f_{(a+2U)}$ equals f_a when U is odd, and $f_{(a+U)}$ equals f_a when U is even, and similarly for g_a.

It can be shown also that when $(\beta_1/R_1) = (\beta_2/R_2) = \ldots = (\beta/R)$, the following conditions apply. When U is odd, all the g_a's are zero, and of the f_a's only those remain for which a is zero or an even multiple of U. When U is even, all the g_a's are again zero, and of the f_a's, only those remain for which a is zero or any multiple of U. Furthermore, in both cases, each of the remaining coefficients is equal to $U\beta/R$.

Thus when U is odd,

$$T = K_v T_t U(\beta/R)(A_0 + A_{2U} \cos 2U\alpha + A_{4U} \cos 4U\alpha + \ldots) \; . \; (15a)$$

and when U is even,

$$T = K_v T_t U(\beta/R)(A_0 + A_U \cos U\alpha + A_{2U} \cos 2U\alpha + \ldots) \; . \; (15b)$$

Cases may also arise where not all the secondary circuits of a particular phase system are in operation. An example of this is given in Section 4.

(3.1) A Practical Expression for Torque

When the coefficients f_a and g_a in eqn. (14) are not zero, the extent of each space harmonic of torque will be affected by the corresponding value of A_a, which in turn depends on the k_n's and $\bar{\sigma}_n$'s. By definition, $k_1 = 1 = \bar{\sigma}_1$, and in order to use this fact to determine the relative values of the A_a's, eqns. (9) and (11) are written so that all terms involving k_1 and $\bar{\sigma}_1$ appear extracted from the various summations.

Thus from eqn. (9)
$$A_0 = 1 + \sum_{n=3}^{\propto} \bar{\sigma}_n k_n^2 \quad . \quad . \quad . \quad . \quad . \quad (16)$$

and from eqn. (11)

$$A_2 = \bar{\sigma}_3 k_3 k_1 + \sum_{n=5}^{\propto} \bar{\sigma}_n k_n k_{(n-2)} + \bar{\sigma}_1 k_1 k_3 + \sum_{n=3}^{\propto} \bar{\sigma}_n k_n k_{(n+2)} - 1$$
$$= (1 + \bar{\sigma}_3)k_3 - 1 + \sum_{n=3}^{\propto} [\bar{\sigma}_n + \bar{\sigma}_{(n+2)}] k_n k_{(n+2)} \quad . \quad . \quad . \quad (17)$$

$$A_4 = (1 + \bar{\sigma}_5)k_5 - (1 + \bar{\sigma}_3)k_3 + \sum_{n=3}^{\propto} [\bar{\sigma}_n + \bar{\sigma}_{(n+4)}] k_n k_{(n+4)} \; . \; (18)$$

$$A_{(a>4)} = [1 + \bar{\sigma}_{(a+1)}] k_{(a+1)} - [1 + \bar{\sigma}_{(a-1)}] k_{(a-1)}$$
$$+ \sum_{n=3}^{\propto} [\bar{\sigma}_n + \bar{\sigma}_{(n+a)}] k_n k_{(n+a)} - \sum_{n=3}^{a-3} \bar{\sigma}_n k_n k_{(a-n)} \quad . \quad (19)$$

It is clear from eqn. (2) that k_n is the ratio of the nth space harmonic of brush voltage to the fundamental, whether the voltage be measured in relation to either β or α. For a typical machine k_n is unlikely to be more than a few per cent at the most, so that all products such as $k_n k_m$ are negligible compared with the single coefficients k_n or k_m. Again, apart from $\bar{\sigma}_1$, which is unity, the ratios $\bar{\sigma}_n$ are all less than unity, and for low values of n they are small compared with unity. Also, since the magnitudes of the flux harmonics in general diminish as the orders increase, all the terms under the summation signs in eqns. (16) to (19) can be neglected.

Thus
$$A_0 = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$
$$A_2 = [(1 + \bar{\sigma}_3)k_3 - 1] \quad . \quad . \quad . \quad . \quad (21)$$
$$A_{(a>2)} = \{[1 + \bar{\sigma}_{(a+1)}] k_{(a+1)} - [(1 + \bar{\sigma}_{(a-1)}] k_{(a-1)}\} \; . \; (22)$$

With the assumptions made, these equations show that the ath torque harmonic depends entirely on the $(a - 1)$th and the $(a + 1)$th space harmonic of flux.

Neglecting flux harmonics of orders higher than five, eqn. (22) shows that the total torque expression will not include torque harmonics of orders higher than six. Eqn. (14) then reduces to

$$T = K_v T_t [f_0 A_0 + (f_2^2 + g_2^2)^{1/2} A_2 \cos (2\alpha + \eta_2)$$
$$+ (f_4^2 + g_4^2)^{1/2} A_4 \cos (4\alpha + \eta_4)$$
$$+ (f_6^2 + g_6^2)^{1/2} A_6 \cos (6\alpha + \eta_6)] \quad . \quad (23)$$

where $\eta_a = \text{arc tan}\,(g_a/f_a)$.

This is a fair representation of the torque of the Schrage motor operating at synchronous speed.

In certain machines with open stator slots, however, higher-order flux harmonics due to slot openings cannot be neglected, and may result in a torque harmonic not included in eqn. (23).

It is seen that the magnitude of each torque harmonic depends both on asymmetrical secondary circuits and on the presence of flux harmonics. For this reason, such a general result as that given by eqn. (23) could not be obtained by dealing separately with the asymmetry and flux harmonics. The rather complicated analysis given above is therefore necessary in order to give a reasonably true representation of the torque at synchronous speed.

The torque expression does, of course, take a simpler form when applied to a particular machine and when the number of secondary phases is known. For example, in a machine with a 3-phase secondary system

$$(f_2^2 + g_2^2) = (f_4^2 + g_4^2) = f^2$$
$$\eta_4 = -\eta_2; \quad f_6 = f_0; \quad \text{and} \quad g_6 = \eta_6 = 0$$

so that the torque equation becomes

$$T = K_v T_t [f_0 A_0 + f A_2 \cos (2\alpha + \eta_2)$$
$$+ f A_4 \cos (4\alpha - \eta_2) + f_0 A_6 \cos 6\alpha] \quad . \quad (24$$

The case of a so-called 2-phase secondary system is dealt with ir Sections 4 and 12.

(3.2) The Torque Components due to Flux Harmonics

The expressions for the A_a coefficients are seen to depenc entirely on the flux harmonics, except for A_0 and A_2, both o which are approximately unity. The coefficient A_0 corresponds to the steady torque, but A_2 applies to the second harmonic of torque. Thus, even when the air-gap flux distribution is purely sinusoidal, the second torque harmonic still remains. Therefore, if the torque harmonics due to non-sinusoidal flux distribution only are to be investigated, it is advisable to eliminate the term $\cos(2\alpha + \eta_2)$.

The only way of accomplishing this is to make all the brush-spread/secondary-resistance ratios equal. The result of making this adjustment is shown in eqns. (15*a*) and (15*b*), where it is seen that a number of harmonics other than the second are also eliminated. As a result, the torque harmonic remaining in a machine with, say, five or more secondary circuits, would probably be extremely weak. For example, for $U = 5$ the lowest torque harmonic remaining is the tenth [see eqn. 15*a*)], and from eqn. (22) it is seen that its magnitude depends on the ninth and eleventh space harmonic of flux.

With either three or six secondary circuits, the sixth torque harmonic would remain; this depends on the fifth and seventh flux harmonics.

Thus, in order to investigate torque harmonics resulting from the presence of flux harmonics, it would be better to work on a machine with a low even number of secondary circuits. For example, when $U = 4$, the fourth torque harmonic remains; this is also true in the special case where only two of the four secondary circuits are utilized. The remainder of the paper is devoted to this latter case, and amply illustrates the application of the foregoing analysis.

(4) MACHINE WITH TWO SECONDARY CIRCUITS

The Schrage motor investigated by the author, and for which test results are given in Section 7, is a machine with a so-called 2-phase secondary system. The relative electrical space displacement of the two secondary phases is, however, $\frac{1}{2}\pi$ and not π as would be required for a bi-phase system. Thus the secondary system actually consists of two consecutive phases of a 4-phase system. However, two equal voltages in quadrature are generally regarded as constituting a 2-phase system, so that this description is retained in the paper. The whole question arises when applying the foregoing equations to the 2-phase case, as will be seen presently.

By amending the torque equation already derived to apply in particular to the 2-phase case, test and theoretical results may be compared. This special consideration is further justified in view of the fact that there is a growing tendency to design small and medium-sized Schrage motors with 2-phase secondary circuits. Such a design gives a more economical machine in that there is a reduction in brush gear; there is also more scope for using split stator windings, which give more versatility in performance.

Although the discussions which form the larger part of the remainder of the paper apply in particular to the case considered, they are, however, similar to those which would apply to any other form of Schrage machine.

For a complete 4-phase secondary case, the particular values of f_0, f_a, and g_a obtained from eqns. (8), (12) and (13) are

$$f_0 = \left(\frac{\beta_1}{R_1} + \frac{\beta_2}{R_2} + \frac{\beta_3}{R_3} + \frac{\beta_4}{R_4}\right)$$

$$f_a = \left[\frac{\beta_1}{R_1} + (-1)^{a/2}\frac{\beta_2}{R_2} + \frac{\beta_3}{R_3} + (-1)^{a/2}\frac{\beta_4}{R_4}\right]; \quad g_a = 0$$

As has been shown, this 4-phase system becomes the 2-phase system when only two consecutive phases are made operative. This means that R_3 and R_4, say, become infinite, so that

$$f_0 = \left(\frac{\beta_1}{R_1} + \frac{\beta_2}{R_2}\right); \quad f_a = \left[\frac{\beta_1}{R_1} + (-1)^{a/2}\frac{\beta_2}{R_2}\right]; \quad g_a = 0$$

Thus in the 2-phase case the torque, from eqn. (14), becomes

$$T = K_v T_t\left[\left(\frac{\beta_1}{R_1} + \frac{\beta_2}{R_2}\right)(A_0 + A_4\cos 4\alpha + A_8\cos 8\alpha + \ldots) + \left(\frac{\beta_1}{R_1} - \frac{\beta_2}{R_2}\right)(A_2\cos 2\alpha + A_6\cos 6\alpha + \ldots)\right] \quad . \quad (26)$$

In a practical case, eqns. (20)–(23) can be used to evaluate the A_a's, and if the effect of flux harmonics of orders higher than five are neglected, the torque harmonics will be limited to those of orders less than eight. Eqn. (26) then reduces to

$$T = K_v T_t\left[\left(\frac{\beta_1}{R_1} + \frac{\beta_2}{R_2}\right)(A_0 + A_4\cos 4\alpha) + \left(\frac{B_1}{R_1} - \frac{\beta_2}{R_2}\right)(A_2\cos 2\alpha + A_6\cos 6\alpha)\right]$$

or, writing s for $(\beta_1/R_1) + (\beta_2/R_2)$, and d for $(\beta_1/R_1) - (\beta_2/R_2)$,

$$T = K_v T_t(sA_0 + dA_2\cos 2\alpha + sA_4\cos 4\alpha + dA_6\cos 6\alpha) \quad . \quad (27)$$

This equation now represents a torque characteristic that may well be obtained in practice. As already shown, if the torque variations due to flux harmonics only are to be investigated, it is as well to eliminate the term involving $\cos 2\alpha$. It is clear from eqn. (26) that this can be accomplished by equating β_1/R_1 and β_2/R_2, thus leaving only the steady component and torque harmonics of orders 4, 8, . . .; i.e. when $\beta_1/R_1 = \beta_2/R_2$,

$$T = K_v T_t\left(\frac{\beta_1}{R_1} + \frac{\beta_2}{R_2}\right)(A_0 + A_4\cos 4\alpha + A_8\cos 8\alpha + \ldots) \quad . \quad (28)$$

There seems no doubt that the $\cos 4\alpha$ term in this equation is responsible for the quadruple slip-frequency speed hunting mentioned in Section 1.

The extent of the 2*a*th harmonic is expressed by the ratio A_{2a}/A_0, which on using eqns. (20) and (22) becomes

$$(A_{2a}/A_0) = \left\{[1 + \bar{\sigma}_{(2a+1)}]k_{(2a+1)} - [1 + \bar{\sigma}_{(2a-1)}]k_{(2a-1)}\right\} \quad . \quad (29)$$

In particular, for the fourth torque harmonic,

$$(A_4/A_0) = [(1 + \bar{\sigma}_5)k_5 - (1 + \bar{\sigma}_3)k_3] \quad . \quad . \quad (30)$$

The test results given to support the present theory concentrate on determining this ratio.

(5) EXPERIMENTAL DETERMINATION OF THE TORQUE CHARACTERISTIC

(5.1) Method of Making Measurements at Synchronous Speed

The Schrage motor is essentially an asynchronous machine, so that in order to make reliable measurements at synchronous speed, it must be forced to run synchronously. The obvious way of accomplishing this is to couple the Schrage motor mechanically to a synchronous machine having the same synchronous speed. The synchronous position of the common shaft will then depend on (*a*) the load on the shaft, and (*b*) the phase relationship between the voltage applied to the synchronous machine and that applied to the Schrage motor, both voltages being of mains frequency. Thus, if a phase shifter is introduced between common supply terminals and the synchronous machine, the synchronous position of the shaft will depend on the phase-

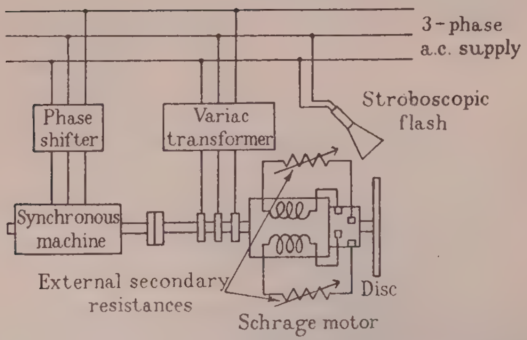


Fig. 3.—Circuit used for testing.

shifter setting. This system was used for experimental investigation, the circuit being shown diagrammatically in Fig. 3. In order to obtain the torque characteristic suggested in eqn. (27), it is necessary to measure the torque (or a known function of it) and the angle α. The torque, measured in synchronous watts, is equal to the input power to the primary winding of the Schrage motor less the appropriate losses (see Section 5.2.3). Thus, provided that these losses can be estimated, the torque can be obtained by measuring the input power.

Provided that the angle of lag between the primary m.m.f. and flux axes remains small or constant, the angle α can be measured by observing the synchronous position of the rotor. In the present case, this was done by the well-known stroboscopic flash method.

(5.2) Practical Details

(5.2.1) External Resistances.

Since the machine under test was of the 2-phase secondary type, only two of the ratios β/R had to be considered. In order to have separate control over β_1/R_1 and β_2/R_2, variable resistors were included in the secondary circuits 1 and 2. These provided the means of increasing the difference between β_1/R_1 and β_2/R_2 inherent in the machine, which thus accentuated some of the oscillations represented in eqn. (27). Also the external resistances made it possible to equate β_1/R_1 and β_2/R_2, so that eqn. (28) could be checked.

The resistances R_1 and R_2 include what is commonly known as "the brush contact resistance," which depends on the current passing through it. Strictly, $(\beta_1/R_1) + (\beta_2/R_2)$, and more particularly $(\beta_1/R_1) + (\beta_2/R_2)$, are functions of secondary current, and are thus functions of torque. Therefore this effect must be made negligibly small, or the foregoing equations must be modified to take it into account. The former end was, in fact, approached by means of the external secondary resistances, which ensured that the purely resistive part of the total resistance of each circuit was much greater than the contact resistance.

Variation in the torque due to brush-contact resistance is discussed further in Section 8.1, in which the practical torque equation already derived is modified.

(5.2.2) Variation of α.

This was done by adjustment of the phase shifter, which in the author's tests took the form of a 3-phase induction motor. To minimize the effect of hysteresis lag between the m.m.f. and flux in the air-gap of the Schrage motor, all readings were taken in a continuous manner, with the phase-shifter adjustment after each reading made as gently as possible and always in the same direction for one complete set of results. The angle α was varied over an interval of 2π, which thus provided a means of investigating the possible presence of reluctance torques due to eccentricity of the rotor or non-uniformity in the stator iron.

(5.2.3) Losses.

For a Schrage motor operating at constant speed the gross output power measured on a different scale gives the torque. The gross output power can in turn be obtained from the input power by subtracting all those losses which the input can supply without magnetically crossing the air-gap. When the speed is synchronous, these losses include the total copper and iron losses. Since the present tests involve measuring torque in terms of input power, it was necessary to determine these losses.

The copper losses were not obtained by the usual short-circuit test, in which the rotor is locked, because the conditions of such a test would differ from the running conditions in that the currents in the stator windings would not be direct currents, and the commutation losses would not be completely represented. The secondary copper loss was, in fact, measured in the following way: The machine was driven at synchronous speed with no primary supply but with direct current supplied from an external source to either of the secondary circuits. The d.c. power required to produce a secondary current equal to the maximum obtained in each of the main tests then gave the corresponding secondary copper loss. The primary copper loss was included with the iron losses obtained by the test described below.

The iron losses at synchronous speed were obtained by driving the Schrage motor at this speed with the secondary windings disconnected (brushes remaining in position) and measuring the primary input power at the appropriate voltage. This power measurement gave the sum of the primary i^2R loss, the primary iron losses, and the power required to produce hysteretic torque. The latter power component in such a test can take any value between ±(secondary hysteresis loss at standstill), depending on the relative position of the rotor and and stator pole systems. It was found, however, that, by changing the flux-axis position α carefully and always in the same direction, the total power input on open-circuit remained constant and independent of α. The possibility of the hysteretic torque varying was therefore eliminated, and the total power input was assumed to represent the primary i^2R loss and total iron losses present in the main torque determining tests.

(6) THE DETERMINATION OF COEFFICIENTS k_n

(6.1) At Synchronous Speed

The coefficients A_a can be obtained indirectly by determining the values of k_n and $\bar{\sigma}_n$. The latter involve only the winding constants of the machine (including brush spread), while a harmonic analysis of a brush-voltage/flux-axis position curve will yield the former [if it is assumed that eqn. (1) represents also the air-gap flux with the secondary windings on open-circuit].

The brush voltage of the secondary circuit 1 can be obtained by summating the expression for V_{un} in eqn. (2) and putting $u = 1$.

$$\text{Thus} \quad V_1 = \sum_{n=1}^{\infty} K_n \sin n\alpha, \quad \text{where} \quad K_n = 2K_v k_n \sin \tfrac{1}{2}\beta_1$$

and neglecting harmonics of orders higher than seven,

$$V_1 = K_1 \sin \alpha + K_3 \sin 3\alpha + K_5 \sin 5\alpha + K_7 \sin 7\alpha \quad . \quad (31)$$

If the Schrage motor is driven at synchronous speed, with the primary winding excited but with the secondary windings on open-circuit, the relationship between brush voltage and main flux position with constant brush spread can be obtained. Eqn. (31) will adequately represent such a curve. The K_n's can be obtained from the following relationships:

$$K_1 = \tfrac{1}{3}[V_{90} + \sqrt{(3)}V_{60} + V_{30}]; \quad K_3 = \tfrac{1}{3}(2V_{30} - V_{90});$$

$$K_5 = \tfrac{1}{6}[V_{90} + 4V_{30} - 3\sqrt{(2)}V_{45}];$$

$$K_7 = \tfrac{1}{6}\{2[\sqrt{(3)}V_{60} + V_{30}] - [V_{90} + 3\sqrt{(2)}V_{45}]\};$$

where V_{30}, V_{45}, V_{60} and V_{90} are the voltages when $\alpha = 30°$, 45°, 60° and 90° respectively, and are best obtained from a graph of V_1 versus α.

The k_n's are then given by $k_n = K_n/K_1$ and can be used directly to find the extent of the space harmonics of torque for the machine operating with the same brush separation as that used in the determination of the K_n's. However, if the brush separations are different for the two cases, the following relationship can be used:

$$\bar{k}_n = k_n(\sin \tfrac{1}{2}\beta \sin \tfrac{1}{2}n\bar{\beta})/(\sin \tfrac{1}{2}\bar{\beta} \sin \tfrac{1}{2}n\beta)$$

where the $\bar{k}_n$'s are the coefficients for brush separation $\bar{\beta}$ and the k_n's are for brush separation β.

(6.2) At Standstill

Eqn. (1) represents the steady air-gap flux density at synchronous speed. At any speed other than synchronous, and in particular at standstill, this flux will become a travelling or rotating wave. Thus, with respect to fixed points in the air-gap, the fundamental and harmonics will then become time variants. Consequently, the voltage appearing at any brush pair will, in general, consist of a fundamental and odd time harmonics. With the use of a suitable waveform analyser, each time harmonic can be picked out and measured. The most convenient condition is that at standstill (and open-circuit secondary windings), when the brushes can be dispensed with and the voltage can be measured directly at the commutator bars. This offers a second method of measuring the k_n's, provided that a relationship between conditions at synchronism and standstill can be obtained.

As pointed out in Section 3.1, the seventh- and higher-order harmonics have negligible effects, so that the fundamental, third and fifth harmonics only need be considered. [The seventh harmonic is included in eqn. (31) merely for the purposes of analysing the voltage wave.] Thus, from eqn. (1) the air-gap flux density at synchronism reduces to

$$B = B_1 \cos \theta - B_3 \cos 3\theta + B_5 \cos 5\theta \quad . \quad . \quad (32)$$

If θ is replaced by θ', where θ' is measured from a datum radius fixed relative to the stator, eqn. (32) will also represent flux conditions at standstill, when the now rotating axis of maximum flux density is instantaneously coincident with the fixed datum.

At t seconds later, when the flux wave has moved in, say, a clockwise direction through an angle ωt radians, the flux-density wave is given by

$$B = B_1 \cos (\theta' - \omega t) - B_3 \cos 3(\theta' - \omega t) + B_5 \cos 5(\theta' - \omega t) \quad . \quad . \quad . \quad . \quad . \quad (33)$$

This is, however, not a complete representation of the flux-density wave at standstill, because eqn. (1) itself is not complete. Eqn. (1), in fact, excludes all flux components at synchronous speed which may be rotating, because they produce torque effects which are negligible. However, the only rotating-flux-density waves unaccounted for at synchronism, which would affect any one of the three components included in eqn. (22), are those which at standstill would rotate in an anti-clockwise direction at speeds of ω, 3ω and 5ω radians per second. These would combine respectively with the components of eqn. (22) to give pulsating flux-density waves. In general, if a component of flux density $B_n \cos n\theta'$, which is stationary at synchronous speed, corresponds to a pulsating flux-density wave at standstill, the pulsating wave is given by

$$[B_n \cos (n\theta' - n\omega t) + B_n \cos (n\theta' + n\omega t)]$$

which represents the clockwise and anti-clockwise rotating flux-density waves. The combination is then $2B_n \cos n\theta' \cos n\omega t$.

In the machine investigated by the author, it was found that in the flux produced by the primary winding acting alone at standstill the only pulsating wave was that of the third time harmonic (harmonics of orders higher than seven being undetected). In fact, for any 3-phase machine of the induction-motor type, if a resultant third harmonic of air-gap flux density exists, it will in general constitute a pulsating wave at standstill.

Thus the relevant part of the main air-gap flux density produced by the primary windings at standstill is given by

$$\begin{aligned} B = B_1 \cos (\theta' - \omega t) - B_3 \cos (3\theta' - 3\omega t) \\ - B_3 \cos (3\theta' + 3\omega t) + B_5 \cos (5\theta' - 5\omega t) \quad . \quad . \quad (34) \end{aligned}$$

$$\begin{aligned} = B_1 \cos (\theta' - \omega t) - 2B_3 \cos 3\theta' \cos 3\omega t \\ + B_5 \cos (5\theta' - 5\omega t) \quad . \quad (35) \end{aligned}$$

It must be emphasized that this does not represent the entire air-gap flux density at standstill, but only the components which become motionless at synchronous speed.

Let us now consider the voltage produced at the brushes by each of the flux-density components represented in eqn. (34). For any component $B_n \cos (n\theta' - m\omega t)$, the voltage at a brush pair with a spread of β and their axis at an angle α' from the datum radius is

$$V'_n = -\frac{m}{n}2K_v k_n \sin \tfrac{1}{2}\beta \sin (n\alpha' - m\omega t)$$

$$= -\frac{m}{n}2K_v k''_n \sin \tfrac{1}{2}n\beta \sin (n\alpha' - m\omega t)$$

where $$k''_n = k_n(\sin \tfrac{1}{2}\beta)/(\sin \tfrac{1}{2}n\beta) \quad . \quad . \quad . \quad . \quad (36)$$

and is in fact independent of β. The other symbols have been previously defined. The instantaneous fundamental, third and fifth harmonics are then, respectively,

$$V'_1 = -2K_v k''_1 \sin \tfrac{1}{2}\beta \sin (\alpha' - \omega t) \quad . \quad . \quad (37)$$

$$\begin{aligned} V'_3 &= 2K_v k''_3 \sin \tfrac{3}{2}\beta[\sin (3\alpha' - 3\omega t) - \sin (3\alpha' + 3\omega t)] \\ &= -4K_v k''_3 \sin \tfrac{3}{2}\beta \cos 3\alpha' \sin 3\omega t \quad . \quad (38) \end{aligned}$$

$$V'_5 = -2K_v k''_5 \sin \tfrac{5}{2}\beta \sin (5\alpha' - 5\omega t) \quad . \quad . \quad (39)$$

From these equations it is seen that the magnitudes of the fundamental and the fifth time harmonic are independent of α but vary with β. Provided that 3α is made equal to any odd multiple of $\tfrac{1}{2}\pi$, all three voltages will have similar form, and their maximum values will be given by the following equations:

$$V'_{1\,max} = 2K_v k''_1; \; V'_{3\,max} = 4K_v k''_3; \; V'_{5\,max} = 2K_v k''_5$$

Since

$k''_1 = 1$, these give $k''_3 = V'_{3max}/2V'_{1\,max}$ and $k''_5 = V'_{5\,max}/V'_{1\,max}$

Thus, k_3 and k_5 for any desired value of β can be obtained by evaluating k''_3 and k''_5 in turn in eqn. (36).

(6.2.1) Experimental Requirements.

To determine the coefficients k''_3 and k''_5 in practice, it is necessary to measure the fundamental, third and fifth time harmonics of brush voltage at standstill for various values of the brush spread β, and to plot the resulting relationship. For the fundamental and fifth harmonic it is immaterial where the brush axis (given by α') lies, but for the third harmonic it is desirable to find the brush position which gives maximum voltage for some suitable constant value of β. Since the rotor is at rest, the brushes can, in fact, be removed and voltmeter prods used, provided that their positioning simulates normal brush movements. The space relationship of each time harmonic of brush voltage at standstill will not in general be represented entirely by eqns. (37), (38) or (39) but will show various other space harmonics. From graphs of magnitudes of the time harmonic voltages against β, it is therefore necessary to extract the relevant space harmonic. Thus the fundamental space harmonic is required from the fundamental time-harmonic curve; the third space harmonic from the third time-harmonic curve; and the fifth space harmonic from the fifth time-harmonic curve. Harmonics other than the fundamental and the third, respectively, in the first and second of the above three cases, were found to be negligible for the particular machine investigated; thus this provided no complication.

In the third case, however, a fundamental space harmonic greater than the fifth itself was obtained, so that graphical harmonic analysis of the usual nature was necessary in order to extract the fifth space harmonic.

(7) EXPERIMENTAL RESULTS

The graphical and tabular results given in this Section are for the 2-phase Schrage motor already mentioned. Figs. 4–7 show the variation of developed torque and secondary currents with the synchronous position of the rotor under different conditions. For the curves of Fig. 4, external resistance was introduced into the secondary circuit 2 to simulate asymmetry. It is seen in this Figure that disturbances in the sinusoidal space

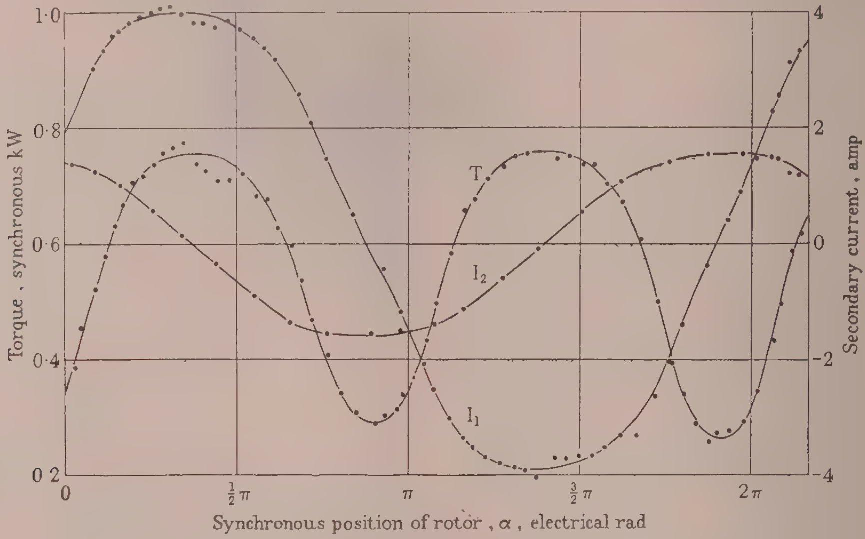


Fig. 4.—Variation of torque and secondary currents with synchronous position.

T = Torque. I_1 = Current in secondary circuit 1.
I_2 = Current in secondary circuit 2.
Applied primary voltage = 250 volts.
Brush spread, β = 1·5 electrical deg.
Ohmic resistance of secondary circuit 1 = 0·286 ohm.
Ohmic resistance of secondary circuit 2 = 0·767 ohm.

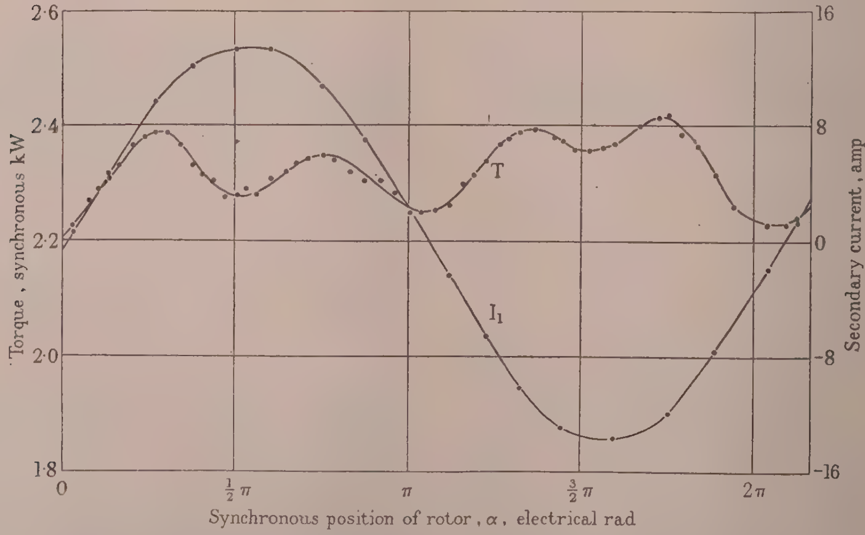


Fig. 5.—Variation of torque and secondary currents with synchronous position.

T = Torque. I_1 = Current in secondary circuit 1.
Applied primary voltage = 220 volts.
Brush spread, β = 30 electrical deg.
Ohmic resistance of secondary circuit 1 = 0·730 ohm.
Ohmic resistance of secondary circuit 2 = 0·714 ohm.

ariations of the secondary currents (noticeable at the peaks of urrent in secondary phase 1) produce proportionate disturbances n the torque curve. It was possible to avoid these disturbances n subsequent tests by always having external resistances present n both secondary circuits. This caused the effective brush ontact resistance, which is erratic in nature, to be swamped, which thus stabilized the total circuit resistances.

The curves of Figs. 5, 6 and 7 were taken for different values of applied primary voltage, while in each case the external esistances were adjusted to make the secondary circuits ymmetrical.

The extent of torque variation in each case is given by the atio of the amplitude of oscillations to the average value; i.e. the ratio $(T_{max} - T_{min})/(T_{max} + T_{min})$. This ratio for each case is given as a percentage in column 3 of Table 2.

From the foregoing theory, the torque represented in Fig. 4 is given by eqn. (27). The ratio $(T_{max} - T_{min})/(T_{max} + T_{min})$ can be obtained from this equation; if we assume A_6 to be negligible, its value is $dA_2/s(A_0 + A_4)$. From eqn. (3) it can be shown that $d/s = (I_1 - I_2)/(I_1 + I_2)$, where I_1 and I_2 are the maximum values of the secondary currents.

The torque represented in each of Figs. 5, 6 and 7 is given by

$$T = sK_vT_t(A_0 + A_4 \cos 4\alpha)$$

and the torque variation becomes A_4/A_0, which is the extent of the fourth torque harmonic. To obtain A_2 and A_4 $(A_0 = 1)$ for

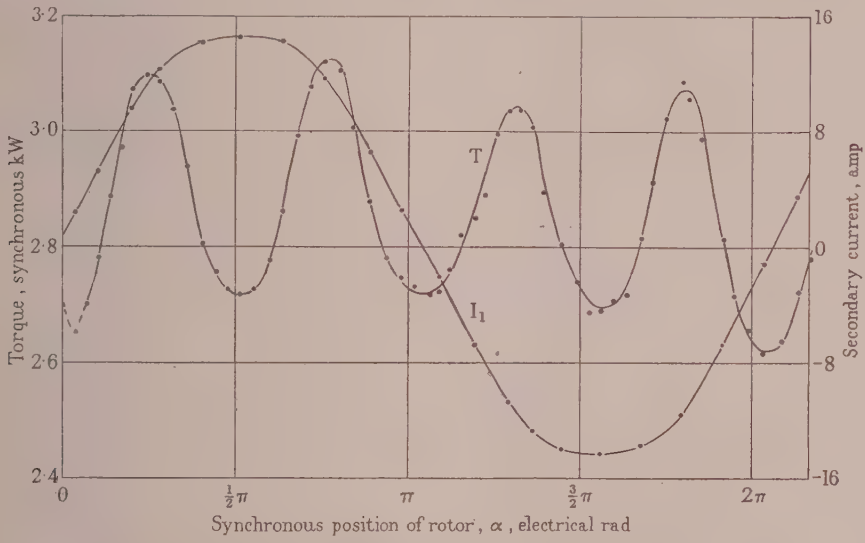


Fig. 6.—Variation of torque and secondary currents with synchronous position.

T = Torque. I_1 = Current in secondary circuit 1.
Applied primary voltage = 250 volts.
Brush spread, β = 30 electrical deg.
Ohmic resistance of secondary circuit 1 = 0·730 ohm.
Ohmic resistance of secondary circuit 2 = 0·714 ohm.

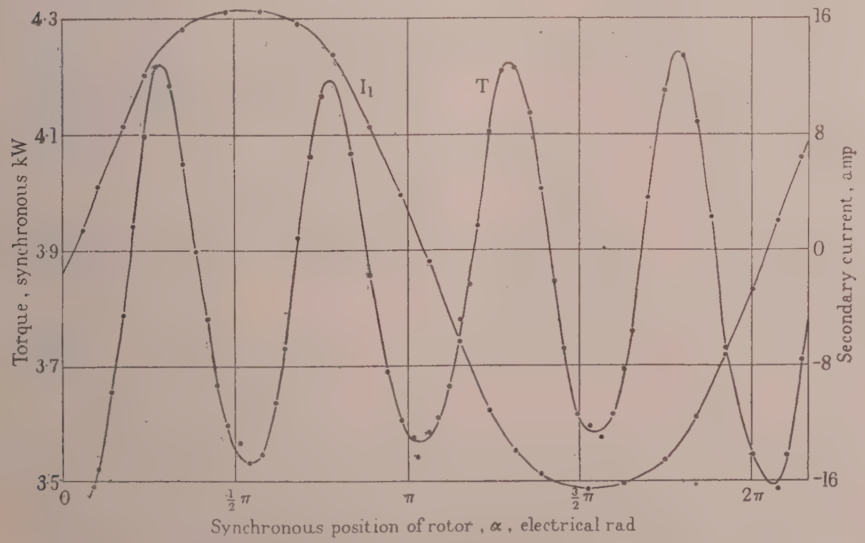


Fig. 7.—Variation of torque and secondary currents with synchronous position.

T = Torque. I_1 = Current in secondary circuit 1.
Applied primary voltage = 280 volts.
Brush spread, β = 30 electrical deg.
Ohmic resistance of secondary circuit 1 = 0·730 ohm.
Ohmic resistance of secondary circuit 2 = 0·714 ohm.

these various curves, the corresponding coefficients k_3 and k_5 are required. These were in turn obtained by the methods described in Sections 6.1 and 6.2, and are shown in Table 1.

Table 1

PERCENTAGE VOLTAGE HARMONICS FOR 30° BRUSH SPREAD

Primary line voltage	220 volts		250 volts		280 volts	
	k_3	k_5	k_3	k_5	k_3	k_5
	%	%	%	%	%	%
Direct voltage, curve (i)	2·4	−1·2	5·4	−1·8	8·1	−0·5
Direct voltage, curve (ii)	2·8	−1·4	6·8	−1·6	9·0	—
Direct voltage, curve (iii)	3·2	−1·2	6·0	−1·7	7·7	−1·0
Direct voltage, curve (iv)	2·8	−1·3	5·1	−1·6	8·0	−0·5
Direct voltage, average curve	2·8	−1·3	5·8	−1·7	8·2	−0·6
Alternating voltage curves (standstill)	2·8	−1·1	4·8	−0·8	6·7	−0·5

Table 1 shows the result of analysing five direct-voltage curves with the machine running at each of three primary line voltages, together with the results obtained from the standstill measurements. Examples of third and fifth time-harmonic voltage curves obtained in the standstill tests are shown in Fig. 8, where, for the third harmonic, negative values refer to a 180° phase change. (Since the brush-axis position is kept constant throughout, the phase relationship of the brush voltage must be either positive or negative.) The curves drawn actually represent theoretical equations obtained by harmonic analysis of the test results.

Although the values of k_3 and k_5 obtained by any one of the two methods are reasonably consistent, the results of the methods do not compare favourably. It would be difficult to account for this discrepancy without some reliable knowledge of the origin and nature of the flux harmonics. For example, the third time harmonic at standstill is taken to be entirely a pulsating wave, whereas part of it may in fact be rotating.

Since the conditions of the tests described in Section 6.1 were nearer those under which the torque measurements were made, the results of these tests are taken to be more reliable.

From the design data of the machine

$$\bar{\sigma}_3 = 0\cdot172\,(\sin \tfrac{1}{2}\beta)/(\sin \tfrac{3}{2}\beta), \text{ and } \bar{\sigma}_5 = 0\cdot748\,(\sin \tfrac{1}{2}\beta)/(\sin \tfrac{5}{2}\beta)$$

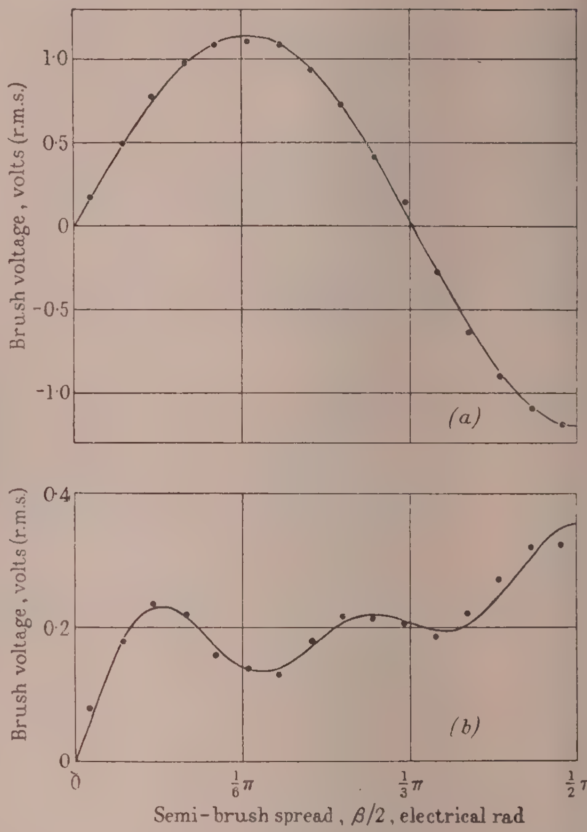


Fig. 8.—Variation of time harmonic voltages with brush spread at standstill.

(*a*) Third time harmonic. The curve is a representation of $V_3 = 1\cdot16 \sin 3(\beta/2) - 0\cdot034 \sin 5(\beta/2)$.

(*b*) Fifth time harmonic. The curve is a representation of $V_5 = 0\cdot285 \sin (\beta/2) + 0\cdot032 \sin 3(\beta/2) + 0\cdot073 \sin 5(\beta/2) + 0\cdot028 \sin 7(\beta/2) + 0\cdot055 \sin 9(\beta/2)$.

With these values, and from Table 1 using values of k_3 and k_5 corresponding to the average direct-voltage curve, suitable substitutions in eqns. (21) and (32) give A_2 and A_4, respectively. Hence predicted values for the percentage torque variations were obtained, and are given in column 4 of Table 2. In column 6 of the same Table corresponding figures are shown calculated by using k_3 and k_5 as obtained by the standstill test of Section 6.2.

Table 2

PERCENTAGE TORQUE VARIATIONS

Condition of secondary windings	Primary line voltage	Brush spread β	Percentage torque variation			
			From torque curves	Predicted by syncronous test	Corrected for contact resistance	Predicted by standstill test (not corrected)
	volts	deg	%	%	%	%
Asymmetrical	250	1·5	44·5	45·8	—	45·2
Symmetrical	220	30	2·3	4·6	2·2	4·3
Symmetrical	250	30	6·7	8·1	5·9	6·1
Symmetrical	280	30	8·4	9·4	7·3	7·7
Symmetrical: Min. external resistances . .	250	12	5·3	9·1	4·3	6·7

Before coming to any conclusions, it is well to consider some obvious sources of error. This is done in the next Section, where the effect of brush-contact drop is shown to be of some importance. In the manner indicated in Section 8.1 corrections were applied to the figures given in column 4 of Table 2, the corrected values being given in column 5.

(8) SOURCES OF ERROR

The errors involved in the present investigation can be divided into two main groups: those arising from actual measurement, and those due to approximations made in the theory.

The first group include errors in obtaining the brush-voltage/flux-axis-position relationship at synchronous speed, meter errors in primary power measurements, and errors in other meter measurements. When fine-grade meters are used, meter errors can be neglected, especially since errors arising from the other sources are usually larger in comparison.

If we consider the brush-voltage curve, the accuracy with which the flux-axis position is determined depends on the secondary hysteresis. Thus should the angle between the rotor synchronous position and the air-gap flux position not be constant while a complete set of readings are taken, the resulting curve would be distorted from its predicted shape. This would mean that the harmonic coefficients obtained by following the analysis of Section 6.1 would be incorrect. Errors of this nature can be minimized by moving the flux axis through a considerable angle before taking any actual readings, while avoiding throughout any reversal in the general direction of flux rotation, and avoiding any jerkiness in the adjustment of the phase shifter.

Oscillations in the rotor synchronous position due to hunting of the synchronous machine are unavoidable, but provided that there is no sudden transient effect, these oscillations are small enough to cause no serious trouble.

The approximations made in the theory include:

(*a*) All errors in winding positions and brush angles (other than a difference in brush spread) are assumed to be negligible.
(*b*) The coefficients $\bar{\sigma}_n$ for the secondary circuit 2 are assumed to be the same as the corresponding coefficients for the secondary circuit 1.
(*c*) Certain assumptions regarding the air-gap flux.
(*d*) Space harmonics of torque of order higher than six are neglected.
(*e*) The total secondary resistances are assumed to be resistive.

By considerably complicating the analysis of Sections 3 and 4 the first and second assumptions were made unnecessary. It was seen that to account for a drop from 8% to, say, 7% in the percentage torque variation due to flux harmonics, angular errors (other than in brush spread) would have to be of the order of 9° electrical (3° mechanical). This conclusion justifies assumption (*a*) for the machine investigated. It was also seen that assumption (*b*) is fully justified, provided that β_1 and β_2 are not large and abnormally unequal.

The assumptions regarding air-gap flux are dealt with in Section 12.

The higher-order torque harmonics are negligible so far as the determination of the relative magnitude of the fourth torque harmonic is concerned. However, as shown in Figs. 6 and 7, they may possibly be sufficient to appear as a distortion in the curve taken to represent a pure fourth space harmonic of torque.

By far the greatest error is that due to (*e*). This will be considered in Section 8.1.

(8.1) The Effect of Brush-Contact Drop

Curve (*a*) in Fig. 9 shows a typical voltage/current characteristic, for increasing and decreasing currents, of an armature circuit consisting of two brush-contact drops and practically negligible ohmic resistance.

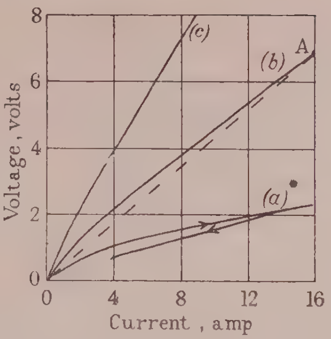


Fig. 9.—Direct-voltage/current characteristic of secondary circuit including brush-contact resistance.

(*a*) Ohmic resistance negligible.
(*b*) Ohmic resistance = 0·285 ohm.
(*c*) Ohmic resistance = 0·722 ohm.

The effective resistance at any point on the curve is obtained by taking the ratio of voltage to current at that point. If the voltage generated in such an armature varies sinusoidally with the position of the axis of a direct exciting flux in the air-gap, the current obtained on short-circuiting the brushes can be computed from the voltage/current characteristic. The relationship between this current and the flux-axis position will, in general, consist of a fundamental and all odd and even harmonics, each having their individual graphical origin. However, when there is some resistance present, as there must be when the brushes are connected to an external winding, the hysteresis effect can be neglected, so that the harmonics then have a common graphical origin. The process is shown in Figs. 9 and 10.

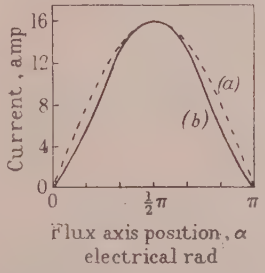


Fig. 10.—Variation of current with flux-axis position.

(*a*) Current with linear resistance and maximum voltage as given by A in Fig. 9.
(*b*) Current corresponding to curve (*b*) in Fig. 9 (difference exaggerated).

To illustrate the application of the above argument, consider the current relationship to contain the fundamental and third harmonic only. Thus, if the brush voltage is given by $V = V_{max} \sin \alpha$, the current on short-circuiting the brushes would be

$$i = I(\sin \alpha - k_b \sin 3\alpha)$$

where k_b is the ratio of peak third harmonic to peak fundamental current. The negative sign is obvious from Figs. 9 and 10.

In eqn. (27) contact resistance affects s and d only, since only these are functions of β_1/R_1 and β_2/R_2. If we consider β_1/R_1, it can be said that $\beta_1/R_1 = i_1\beta_1/V_1$, where V_1 is obtained by applying eqn. (2) to phase 1 for all odd values of m, and i_1 is the current flowing in circuit 1 due to V_1. To simplify the method, let V_1 in the present substitution be $V_1 = V_{1\,max} \sin \alpha$.

Therefore, $i_1 = I_1 (\sin \alpha - k_b \sin 3\alpha)$

and $$\frac{\beta_1}{R_1} = \beta_1 \frac{I_1}{V_{1\,max}} \frac{\sin \alpha - k_b \sin 3\alpha}{\sin \alpha} = \frac{\beta_1}{R_1'}[(1 - k_b) - 2k_b \cos 2\alpha]$$

where $R_1' = V_{1\,max}/I_1$ and is now constant.

Similarly

$$\frac{\beta_2}{R_2} = \beta_2 \frac{I_2}{V_{2\,max}} \frac{\cos \alpha + k_b \cos 3\alpha}{\cos \alpha} = \frac{\beta_2}{R_2'}[(1 - k_b) + 2k_b \cos 2\alpha]$$

where $R_2' = V_{2\,max}/I_2$, which is also constant.

Therefore

$$s = \left(\frac{\beta_1}{R_1} + \frac{\beta_2}{R_2}\right) = \left(\frac{\beta_1}{R_1'} + \frac{\beta_2}{R_2'}\right)(1 - k_b) - \left(\frac{\beta_1}{R_1'} - \frac{\beta_2}{R_2'}\right)$$
$$2k_b \cos 2\alpha = s'(1 - k_b) - 2d'k_b \cos 2\alpha$$

and

$$d = \left(\frac{\beta_1}{R_1} - \frac{\beta_2}{R_2}\right) = \left(\frac{\beta_1}{R_1'} - \frac{\beta_2}{R_2'}\right)(1 - k_b) - \left(\frac{\beta_1}{R_1'} + \frac{\beta_2}{R_2'}\right)$$
$$2k_b \cos 2\alpha = d'(1 - k_b) - 2s'k_b \cos 2\alpha$$

where $$s' = \left(\frac{\beta_1}{R_1'} + \frac{\beta_2}{R_2'}\right) \quad \text{and} \quad d' = \left(\frac{\beta_1}{R_1'} - \frac{\beta_2}{R_2'}\right)$$

Substituting these values for s and d in eqn. (27) and grouping terms, the following expression is obtained:

$$\begin{aligned} T = K_v T_t \{ & s'[A_0 - k_b(A_0 + A_2)] \\ & + d'[A_2 - k_b(2A_0 + A_2 + A_4)] \cos 2\alpha \\ & + s'[A_4 - k_b(A_2 + A_4 + A_6)] \cos 4\alpha \\ & + d'[A_6 - k_b(A_4 + A_6)] \cos 6\alpha - s'k_b A_6 \cos 8\alpha \} \quad . \quad (40) \end{aligned}$$

Eliminating the cos 2α and cos 6α terms as before, but this time making $d' = 0$, and using eqns. (20)–(22) to substitute for the A_a's, and neglecting terms in k_7 and cos 8α,

$$\begin{aligned} T = s'K_v T_t \{ & [1 - (1 + \bar{\sigma}_3)k_3 k_b] \\ & + [(1 + \bar{\sigma}_5)k_5 - (1 + \bar{\sigma}_3)k_3 + k_b] \cos 4\alpha \} \quad . \quad (41) \end{aligned}$$

Since $(1 + \bar{\sigma}_3)k_3 k_b \ll 1$, the extent of the fourth space harmonic of torque now becomes A_4'/A_0, where $A_0 = 1$ and

$$A_4' = [(1 + \bar{\sigma}_5)k_5 - (1 + \bar{\sigma}_3)k_3 + k_b]$$

which is the new coefficient of cos 4α.

Thus, to correct the extent of fourth torque harmonic given by eqn. (30) it is necessary merely to add k_b. In the practical case considered, however, A_4 is negative, so that the effect of non-linear secondary-resistance characteristic is to reduce the effect of the flux harmonics.

Actual measurements of k_b were made with external resistances equal to those used in the main torque-determining tests [see curves (*b*) and (*c*) of Fig. 9]. These are shown as percentages in Table 3 and are applied as corrections in Table 2. The effect of k_b is shown most clearly in a test where symmetrical secondary circuits were produced by the introduction of minimum possible external resistance. (The external resistances given include that of the stator winding.)

Table 3

BRUSH-CONTACT RESISTANCE EFFECT

External resistance	Maximum secondary current	k_b
ohm 0·722	amp 13·4	2·06
0·722	14·2	2·21
0·722	16·4	2·36
0·285 (min.)	16·2	4·80

(9) CONCLUSIONS

The present investigation shows that the variation in the torque developed by a Schrage motor, when produced either by unequal brush spreads or asymmetrical secondary resistances, can be effectively eliminated by the inclusion of suitable resistances in the secondary circuits. However, if the air-gap flux distribution at synchronous speed is not sinusoidal, all the torque variations cannot be eliminated in this way. The mathematical analysis undertaken in Section 3 shows the connection between flux harmonics and torque variations (also referred to as harmonics); the analysis is extended in Section 8.1 to include brush-contact effects. Test results reasonably substantiate the theory (cf. columns 3 and 5 of Table 2).

If in a machine the flux distribution referred to above is flat-topped, and if the brush voltage/current characteristic is similar to that given by curve (*a*) of Fig. 9, the brush-contact effect will partly compensate for the effect of flux harmonics on the torque. Inclusion of external secondary resistances will, however, reduce this compensation (cf. rows 3 and 5 in Table 2).

The simple theory of rotating fields produced by 3-phase windings housed in slots on a rotor shows that, at synchronous speed, stationary space harmonics of flux can exist only when corresponding time harmonics of m.m.f. for each primary phase are present, as well as space harmonics due to winding distribution. Under average conditions, however, stationary flux harmonics produced in this way would be practically negligible; this was, in fact, the case for the machine investigated. Further investigation is therefore suggested in order to establish a theory regarding the origin of the stationary flux harmonics. Also, such a theory would be required to show a closer connection between open-circuit conditions at synchronous speed and at standstill than that suggested in Section 6.2. The test results show that, for the machine tested, the proportion of third harmonic in the flux wave increases with applied primary voltage and therefore with average flux density. This indicates that the flux-harmonic content depends on the degree of magnetic saturation in the machine. Magnetic saturation may therefore play a large part in the production of the flux harmonics which cause the effects analysed and discussed in the paper.

(10) ACKNOWLEDGMENTS

The author would like to thank Prof. J. C. Prescott for encouragement, and Mr. A. O. Carter for valuable advice given in the preparation of the paper.

(11) REFERENCES

(1) ARNOLD, A. H. M.: "The Circle Diagrams of the Three-Phase Shunt Commutator Motor," *Journal I.E.E.*, 1926, **64**, p. 1139.

(2) ADKINS, B., and GIBBS, W. J.: "Polyphase Commutator Machines" (Cambridge University Press, 1951), p. 117.

(3) LIWSCHITZ-GARIK, M., and WHIPPLE, C. C.: "Electric Machinery, Vol. II" (D. van Nostrand, 1947), pp. 236 and 494.

(4) LIWSCHITZ, M. M.: "Field Harmonics in Induction Motors," *Transactions of the American I.E.E.*, 1942, **61**, p. 797.

5) Weber, C. A. M., and Lee, F. W.: "Harmonics due to Slot Openings," *Journal of the American I.E.E.*, 1924, **43**, p. 687.

5) Chapman, F. T.: "Air-Gap Field of an Induction Motor," *Electrician*, 1916, **77**, p. 663.

(12) APPENDIX

Assumptions regarding the air-gap flux can be summarized s follows:

(*a*) The air-gap flux represented by eqn. (1) is assumed to be due to an m.m.f. which is the resultant of primary-, stator- and commutator-winding m.m.f.'s.

(*b*) For constant applied primary voltage, all coefficients in eqn. (1) are assumed to be constant and, in particular, independent of the flux-axis position α.

(*c*) The harmonic ratios of the flux wave, given by B_n/B_1, are assumed to be the same whether the secondary circuits are open or closed.

(*a*) is a definition, and needs no further comment.

The resultant flux depends on the primary voltage less the primary impedance drop, and will be constant in magnitude and form provided that both primary voltage and impedance drop (or primary current) are constant. Usually there is no difficulty in assuming the former to be constant, while in the present case conditions are being investigated when the machine is operating against a torque which varies only between narrow limits; i.e. the primary current remains practically constant (for the test made with balanced secondary circuits, the primary current varied by not more than 3%). Thus, provided that the primary current is constant, the resultant air-gap flux does not depend on either α or the m.m.f. of the secondary circuits (being the only variants), and there is, in fact, justification for assuming it to be constant in magnitude and form.

The assumption made in (*c*) depends on the difference between the primary impedance drop (or primary current) with the secondary circuits open and closed. The resultant flux in most cases will be slightly less on load than on open-circuit. A secondary m.m.f. rich in harmonics would mean more distortion in the primary current (and therefore impedance drop) on load than on open-circuit; this would, in turn, result in the flux harmonic ratios, and therefore k_n, being higher for the resultant flux on load than for that on open-circuit. However, this effect cannot readily be taken into account; it can only be referred to as a possible source of error in the present investigation.

621.313.322:621.3.016.35

Monograph No. 134 S
June 1955

STEADY-STATE STABILITY OF SYNCHRONOUS GENERATORS AS AFFECTED BY REGULATORS AND GOVERNORS

By H. K. MESSERLE, M.Eng.Sc., B.E.E., and R. W. BRUCK, B.Sc., B.E.

(*The paper was first received* 2nd *November*, 1954, *and in revised form* 31*st January*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in June*, 1955.)

SUMMARY

Voltage and power-angle regulation can be used to improve the steady-state stability of synchronous alternators, whereas speed governors and tie-line power controllers often introduce instability.

In the paper, methods of determining the effects of voltage and angle regulation on the steady-state stability limit are discussed, and they are extended to allow for the control of the prime-mover torque by means of governors and power controllers. The machine analysis is based on the general equations for synchronous machines.

A complete analysis of actual problems is rather tedious, and a differential analyser has been found most suitable for detailed investigations. Results, as obtained for typical alternators, are presented in the form of stability contour diagrams, which are very convenient for design purposes, since optimum control parameters can be read off directly.

LIST OF SYMBOLS

(*a*) *For Machine as shown in Fig*. 15.

v_d = Direct-axis voltage.
v_q = Quadrature-axis voltage.
v_t = Generator terminal voltage.
v'_{t0} = Reference voltage, $\Delta v'_t = v_t - v'_{t0}$.
v = Field voltage.
v_{fd} = Equivalent field voltage = $X_{afd}v/R_f$.
i_d = Direct-axis current.
i_q = Quadrature-axis current.
i_f = Field current; $i_{fd} = X_{afd}i_f$.
Φ_d = Direct-axis flux linkage.
Φ_q = Quadrature-axis flux linkage.
Φ_{ffd} = Field flux linkage, and $\Phi_{fd} = X_{afd}\Phi_{ffd}/X_f$.
X_d = Direct-axis synchronous reactance.
X'_d = Direct-axis transient reactance.
X_q = Quadrature-axis synchronous reactance.
X_f = Field-winding reactance.
X_{afd} = Mutual reactance between field and direct-axis armature windings.
R_f = Field-winding resistance.
θ = Displacement of direct axis with respect to stator.
δ = Angle between quadrature axis and infinite bus voltage.
t = Time.
f = Frequency.
ω = Angular velocity.
ω_0 = $2\pi f_0$ × synchronous speed.
f_0 = 50c/s or 1 per unit.
$p = \dfrac{d}{dt}$ = Time derivative.
τ'_{d0} = Open-circuit generator field time-constant.
$\tau'_{dz} = \dfrac{X'_d + X_e}{X_d + X_e}\tau'_{d0}$.
$M = 4\pi f_0 H$.
H = Inertia constant.
D = Damping due to prime mover and load.
T_m = Prime-mover torque.
T_{el} = Electrical torque.
F_c = Transfer function of controller.
F_a = Transfer function of amplifier in regulator.
F_r = Transfer function of exciter and stabilizer.
F_R = Transfer function of regulator.
F_{Mvt} = Transfer function of alternator with voltage regulation.
$F_{M\delta}$ = Transfer function of alternator with angle regulation.
F_{TL} = Transfer function of tie-line power controller.
τ_1, τ_2 = Governor time-constants.
r = Regulation of governor.

(*b*) *For Derivative Transformer.*

v_p = Primary voltage.
i_p = Primary current.
v_c = Secondary voltage.
i_c = Secondary current.
R_p = Primary resistance.
L_p = Primary inductance.
R_c = Secondary resistance.
L_c = Secondary inductance.
N_p = Number of turns on the primary winding.
N_c = Number of turns on the secondary winding.

(*c*) *For Regulator.*

τ_e = Exciter time-constant.
v_b = Stabilizer voltage.
μ_s = Stabilizer gain.
τ_s = Stabilizer time-constant.
v'_a = Amplifier output.
v_a = Amplifier input.
τ_a = Amplifier time-constant.

(1) INTRODUCTION

Steady-state stability generally limits the performance of an alternator when operating at a leading power factor. The operating range throughout the whole leading-power-factor region can be extended considerably with continuous regulation, whereas governors are usually less effective and often reduce the stability limit. The stability limit, together with the general operating limits of an alternator, is of particular importance to the operating engineer. It forms a part of the operating characteristic or capability diagram of the alternator and covers the major part of the leading-power-factor region,[1] as indicated in Fig. 4.

The steady-state stability limit of an alternator defines the maximum steady load the alternator can carry without falling out of synchronism. This limit can be found experimentally by loading the alternator, increasing the load in small steps, until it becomes unstable. Theoretically the load steps should be infinitesimally small to avoid transient disturbances which obscure the results.

When using continuously acting regulators and governors, negative feedback can be introduced, and the alternator may be loaded beyond its ordinary steady-state stability limit and its load angle may increase past the normal maximum. When

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Messerle is at the University of Sydney, New South Wales.
Mr. Bruck is with the Electricity Commission of New South Wales.

pperating beyond the ordinary limit the alternator is said to run n the dynamic stability region, and it falls out of step or becomes nstable when it reaches the "dynamic stability limit."

The dynamic limit depends, not only on the alternator itself, ut also on the load and system characteristics. Usually, however, any particular alternator is only a minor component in a arge supply system. Hence, as a first approximation, the lternator can be considered as being connected to an infinite usbar. Thus the basis for practically all dynamic stability tudies in recent years[1–6] has been the assumption that the lternator is feeding into a relatively large system which can e represented by an infinite busbar as shown in Fig. 1.

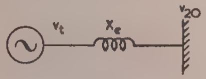


Fig. 1.—Synchronous generator connected to an infinite busbar through a reactance X_e.

The determination of the steady-state stability limit for a nachine with discontinuous regulation is straightforward, and mathematical expression for this limit can be deduced directly rom the steady-state vector diagram.[1] However, when introlucing feedback with regulators or governors, the steady-state ector diagram has to be replaced by a more accurate description f the machine allowing for field time-constant, inertia and ther transient quantities.

The instability of alternators with feedback usually shows up t the dynamic limit in the form of self-excited oscillations in he feedback system and not, as normally happens at the steadytate limit, by a slow falling out of synchronism with a coninuously and monotonically increasing load angle. The period f these oscillations depends on the regulator parameters and anges from about 0·5 to 10sec or more for large machines. Thus a determination of the stability limit by representing a nachine simply by its synchronous reactance does not allow for he fast changes in the variables. Going to the other extreme nd using the transient reactance of the machine is not justified ither, since the transient reactance applies to changes which ake place within a fraction of a second. In spite of these bjections, equivalent reactances for machines simplify the nalysis considerably and have been used in several papers, .g. Reference 7, but an accurate analysis has to be based on he general machine theory as given originally by Park[8] and ound in many textbooks.

For the analysis of the behaviour of the regulator and governor, rdinary servo-mechanism theory applies, and once the complete ystem equations, including those for the alternator, are stablished, the stability limits can be found by using standard tability criteria.[9]

Several papers have discussed the general effect of regulators n the stability limit under particular conditions. By reprecenting the alternator only by an equivalent reactance, Adkins[7] as indicated the use of servo-mechanism theory. Concordia oes a step further,[2,3,4] and using generalized machine theory, e discusses the effect of voltage and angle regulators on the tability limit of round-rotor generators with $X_d = 1\cdot0$ per unit nd operation at unity power factor. He uses the well-known Routh's criteria to derive the stability from the characteristic ystem equation. Heffron and Phillips[5] give a method, also ased on the general machine theory, for the investigation of lternators operating at any power factor, and they establish perating charts for the leading-power-factor region showing ome typical examples and considering voltage regulation.

In this paper the analysis has been extended to allow for the ffects of governors as well as regulators, which means that, not only the control of the alternator field voltage, but also the control of the mechanical input due to speed and tie-line power controllers is considered.

The machine equations used in the paper are derived from the general equations for synchronous machines. They are reduced to a form suitable for stability studies and are similar to the equations used by Heffron and Phillips.[5] They still contain many parameters whose effects are important, and they remain complex in particular when they have to be analysed together with the regulator and governor equations. Thus a detailed investigation becomes very involved unless some automatic computer is used.

In the following Sections, first the machine and regulator equations are established. Then the effects of the important regulator parameters are investigated in detail. Finally, the governor equations are set up and the possible modifications of the steady-state and dynamic stability limits with prime-mover torque control are discussed.

Transfer-function diagrams are used to show graphically the trend in stability when changing the major parameters. For the detailed analysis of actual alternators a differential analyser[10] has been found most suitable, and the method employed when deriving the results, which are presented in this paper, is discussed. Stability contour diagrams (Figs. 7, 8, and 11) are used to record the results in a form showing very clearly the effect of varying regulator parameters.

The analysis covers the case of an alternator feeding into an infinite busbar through an external impedance jX_e. The results obtained for this case are usually sufficient for the operator of an alternator if the supply system itself is large or in particular if the generator is a long distance away from the main system. The approach could be extended to cover the case of several machines operating in parallel or other combinations, although the mathematics involved becomes rather complex.

(2) MACHINE EQUATIONS

The general equations for the synchronous machine as developed by Park[8] have found little application because of their complexity. They are based on three main assumptions:

(*a*) Saturation is neglected.
(*b*) The effect of stator slots is negligible.
(*c*) The stator windings are sinusoidally distributed around the air-gap when considering their interaction with the rotor windings.

Usually additional approximations are made to simplify the resulting equations. In particular, when the steady-state stability is to be determined, the complexity can be reduced by considering only small deviations of the variables from their steady-state values. This, in effect, means that the system equations can be linearized, since only infinitesimally small changes are of interest.

For example, the output voltage v_t of the alternator is then

$$v_t = v_{t0} + \Delta v_t$$

where v_{t0} is the steady-state value before a disturbance occurs, and Δv_t is any small variation in v_t arising when the disturbance is taking place.

Similarly	$\delta = \delta_0 + \Delta\delta$	Load angle
	$v_{fd} = v_{fd0} + \Delta v_{fd}$	Field voltage
	$T_m = T_{m0} + \Delta T_m$	Mechanical torque

where the zero index refers to the undisturbed steady-state value, and Δ to a small change.

If we consider the alternator to be connected to an infinite busbar through an impedance jX_e, the machine equations can be simplified and linearized in terms of the small variations of the variables concerned. The linearized machine equations, as derived in Appendix 9.1, are then

$$[(\tau'_{dz}p + 1)(Mp^2 + Dp + A_1) - A_2]\Delta\delta + A_3\Delta v_{fd} = (\tau'_{dz}p + 1)\Delta T_m \quad . \quad (1)$$

$$(A_4\tau'_{dz}p + A_5)\Delta\delta + A_6\Delta v_{fd} = (\tau'_{dz}p + 1)\Delta v_t \quad . \quad (2)$$

where $M = 4\pi f_0 H$

D = Damping coefficient

$\tau'_{dz} = \dfrac{X'_d + X_e}{X_d + X_e}\tau'_{d0}$

X'_d = Direct-axis transient reactance

$p = \dfrac{d}{dt}$ = Time derivative

H = Inertia constant

f_0 = Frequency, 50c/s

τ'_{d0} = Open-circuit generator field time-constant (per unit)

X_d = Direct-axis synchronous reactance

t = Time (per unit)

Per-unit notation is used throughout the paper unless stated otherwise.

The A_i's are constants and are determined by the steady-state load condition of the system.

The coefficient D is a rather general term allowing, not only for damper windings, but also for prime-mover damping and variations in load torque with changes in speed. In general, D is not a constant and varies with speed and power flow. However, for any particular speed and power when only small changes about a mean value arise, D can be assumed to remain constant. Thus for the determination of the dynamic stability limit the magnitude of D can be established beforehand for any particular condition.

Saturation is another factor that may have to be considered. It can be allowed for by using equivalent values for the machine reactances.[11]

Eqn. (1) can be used for the determination of the steady-state stability limit, i.e. for discontinuous regulation. The field voltage is then to be considered as constant, i.e. $\Delta v_{fd} = 0$, and the mechanical torque variations are zero as well. Thus eqn. (1) reduces to

$$[\tau'_{dz}M_p^3 + (\tau'_{dz}D + M)p^2 + (\tau'_{dz}A_1 + D)p + (A_1 - A_2)]\Delta\delta = 0 \quad . \quad . \quad . \quad . \quad (3)$$

For stability, all coefficients of eqn. (3) must be greater than zero, since otherwise unstable roots would be present. This and Routh's criterion lead to the stability condition

$$A_1 - A_2 \geqslant 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

By using the steady-state vector diagram (Fig. 16) eqn. (4) can be reduced to

$$v_{20}v_{q0}\cos\delta_0 + \frac{v_{20}^2\sin^2\delta_0(X_q - X_d)}{X_e + X_d} \geqslant 0 \quad . \quad . \quad (5)$$

Thus the limit is independent of field time-constant, damping coefficient and transient reactance, and it corresponds to the limit that can be obtained directly from the steady-state vector diagram.[1]

(3) REGULATORS

A regulator compares some machine quantity, such as the output voltage, with a reference quantity, and the difference of the two is used to adjust field voltage and flux. One of its main purposes is to minimize any steady-state and transient variati in the controlled quantity. Since this involves negative feedba it means that the dynamic stability of the machine can also improved in many cases. The dynamic limit is defined here the steady-state limit of an alternator which is operating toget with a continuously acting regulator and/or a governor.

Various quantities may be controlled, such as the out voltage, the power angle or the output current. In general, output voltage can be controlled more readily than the othe It is also possible to achieve by voltage regulation at least same, if not higher, stability than by regulating the ot quantities, as will be shown later.[5]

The last statement does not apply to transient stability large power swings. In any case, the effect of a regulator on transient stability is small in general, because of the relativ long delays involved in the regulator action owing to the m field time-constant.

The stability under steady-state and dynamic conditions is concern in the paper. Usually transient stability limits are considered when determining the operating characteristics of alternator. However, the transient stability limit can beco critical when operating the alternator in the unity-power-fac region[6] if the external system has a high reactance, $X_e > 0{\cdot}5$ per unit.

(3.1) Voltage Regulation

When analysing an alternator with continuous regulation can be considered as a large-scale servo-mechanism and can broken down into a block schematic as shown in Fig. 2. T

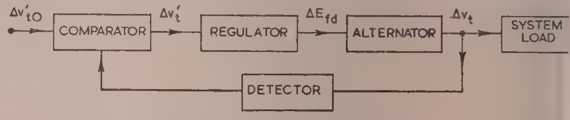


Fig. 2.—Voltage regulation.

alternator is shown with its load, and the output voltag given as

$$v_t = v_{t0} + \Delta v_t.$$

This voltage is compared with the reference

$$v'_{t0} = v_{t0} + \Delta v'_{t0}$$

where $\Delta v'_{t0}$ is the variation in the reference.

The difference

$$\Delta v'_t = v_t - v'_{t0} = \Delta v_t - \Delta v'_{t0} \quad . \quad . \quad . \quad .$$

passes through the controlling amplifier to the exciter an controls the field voltage.

To determine whether the system is stable for any partic load condition, Routh's criterion[9] for stability may be u Additional information can be obtained and the computati work can be speeded up if the open-loop transfer function o system is plotted and Nyquist's criterion[9] or conformal tr

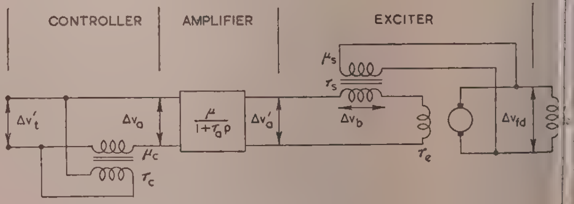


Fig. 3.—A typical regulator circuit.

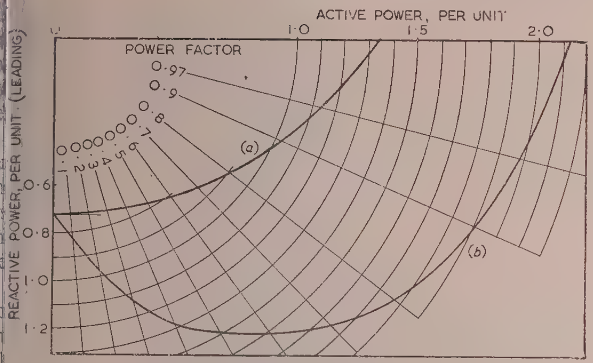


Fig. 4.—Steady-state stability limit of a 30 MW alternator.

(*a*) With discontinuous regulation.
(*b*) With continuous regulation.
$H = 6\cdot0$, $D = 3\cdot0$, $v_{t0} = 1\cdot0$
$X_d = X_q = 1\cdot39$, $X'_d = 0\cdot22$, $X_e = 0\cdot4$
$\tau'_{d0} = 5\cdot8$ sec, $\tau_e = \tau_s = 1\cdot6$ sec, $\mu = 10$, $\mu_s = 3$

ormation is applied. The transfer-function approach has one articular advantage in that each component can be treated eparately and the effect of each on the stability limit can be een directly.[9] Thus the parameters for any particular com-onent may be chosen so that its effect on the whole system is an ptimum.

The open-loop transfer function is

$$F = F_R \times F_{Mvt} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

here F_R = Regulator transfer function
F_{Mvt} = Machine transfer function for voltage regulation.

.1.1) Machine Transfer Function.

The machine transfer function F_{Mvt} can be derived from the achine equations (1) and (2). Thus by eliminating $\Delta\delta$, hese equations lead to

$$_{Mvt} = \frac{\Delta v_t}{\Delta v_{fd}} = \frac{-A_6\left(Mp^2 + Dp + A_1 - \frac{A_2A_4}{1 - A_5}\right)}{M\tau'_{dz}p^3 + (M + D\tau'_{dz})p^2 + (D + A_1\tau'_{dz})p + (A_1 - A_2)} \quad . \quad . \quad . \quad . \quad (8)$$

Ience the machine transfer function depends on the parameters I, D and τ'_{dz}, and also on the initial steady-state load condition s determined by the A$_i$'s.

As an example, a typical transfer function plot is given in ig. 5 for a 30 MW alternator operating at unity power factor with 2·1 times the rated or 2·1 per unit current at rated or 1·0 per unit voltage. The machine, which is used as a typical example throughout this paper, has the following constants:

$$\left.\begin{array}{ll} H = 6\cdot0 \text{ per unit} & X_d = X_q = 1\cdot39 \text{ per unit} \\ \tau'_{d0} = 5\cdot8 \text{sec} & X'_d = 0\cdot22 \text{ per unit} \\ D = 3\cdot0 \text{ per unit} & X_e = 0\cdot4 \text{ per unit} \\ v_{t0} = 1\cdot0 \text{ per unit} & \end{array}\right\} \quad . \quad (9)$$

The transfer function for 2·1 per unit current is given here, since it will be needed later on when the alternator performance at the dynamic stability limit is considered.

(3.1.2) Regulator Transfer Function.

The function F_R for the regulator can be split up into three components:

$$F_R = F_c \times F_a \times F_r \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where F_r = Exciter transfer function
F_n = Amplifier transfer function
F_c = Controller transfer function.

A general regulating system is shown in Fig. 3, which includes a stabilizing transformer. An amplifier, usually of the rotating type such as the amplidyne, is used, but magnetic amplifiers are now coming into use in many modern regulators. The time delay in the amplifier is usually relatively small when compared with that of the other components in the regulating loop of big machines. Thus the amplifier can be represented approximately by a simple transfer function with a single time-constant τ_a. The controller is used to correct the error signal if necessary. The most common method is derivative error control, which is provided here by means of a derivative transformer with gain μ_c and time-constant τ_c.

The transfer function for this regulator as derived in Appendix 9.2 is

$$F_R = \frac{\mu}{(1 + \tau_a p)}\left[1 + \frac{\mu_c\tau_c p}{(1 + \tau_c p)}\right] \left\{\frac{1 + \tau_s p}{\tau_s\tau_e p^2 + [\tau_e + \tau_s(1 + \mu_s)]p + 1}\right\} \quad . \quad (11)$$

This equation is plotted in Fig. 5 for $0 \leqslant p = j\omega \leqslant j\infty$ and assuming that $\tau_a = \mu_c = 0$. The effect of changing the stabilizer gain, μ_s, is shown in the same Figure. As will be seen more clearly later, the curves indicate that an increase in μ_s should make the whole system more stable for high amplifier gains μ, but not for low gains.

(3.1.3) Overall Open-Loop Transfer Function.

The overall transfer function is the product of the machine and regulator functions. If the time delays in the regulator are neglected its transfer function becomes simply

$$F_R = -\mu\,(\mu = |\mu|)$$

which implies a rotation of the phase angle of F_{Mvt} by 180°. For $\mu = 1$ the function has been plotted in Fig. 6(*a*), and it follows that the system is unstable unless the gain is increased to about $\mu = 2\cdot2$. This can be deduced directly from the graph, since μ is a constant multiplier of the overall transfer function, and for stability the zero-frequency point on the curve should lie to the left of the $(-1, 0)$ point on the real axis. When the gain μ is increased to 5·7 the system becomes unstable again, which leads to a stability range

$$2\cdot2 < \mu < 5\cdot7$$

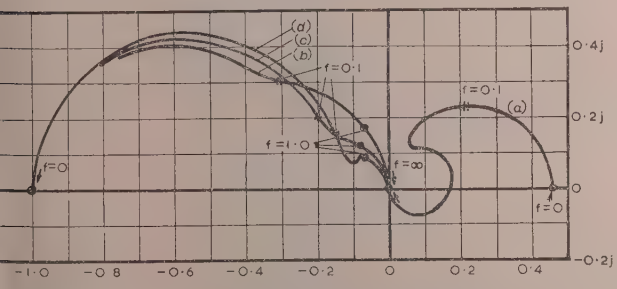


Fig. 5.—Transfer functions for voltage regulation $0 \leqslant f \leqslant \infty$.

(*a*) F_{Mvt} for 1·0 power factor and 2·1 per-unit current.
(*b*) F_R for $\tau_e = \tau_s = 500$, $\mu = 1$, $\mu_s = 1$, $\tau_a = \mu_e = 0$.
(*c*) F_R for $\tau_e = \tau_s = 500$, $\mu = 1$, $\mu_s = 3$, $\tau_a = \mu_c = 0$.
(*d*) F_R for $\tau_e = \tau_s = 500$, $\mu = 1$, $\mu_s = 5$, $\tau_a = \mu_c = 0$.

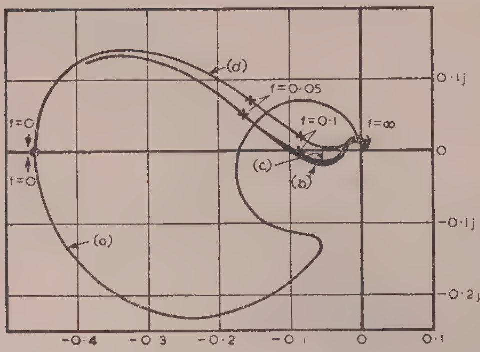


Fig. 6.—Overall transfer function for $\mu = 1$, $\mu_s = 3$, $\mu_c = 0$ and $0 \leqslant f \leqslant \infty$.

(*a*) $\tau_s = \tau_e = 0 = \tau_a$ (ideal voltage regulator).
(*b*) $\tau_s = \tau_e = 500$, $\tau_a = 0$.
(*c*) $\tau_s = \tau_e = 500$, $\tau_a = 30$.
(*d*) $\tau_s = \tau_e = 500$, $\tau_a = 150$.

When the time delays in the regulator are allowed for, the result depends on the actual values of the time-constants τ_e for the exciter and τ_s for the stabilizer. For typical values of these constants, i.e. $\tau_e = 500$ per unit ($= 1\cdot59$ sec) and $\tau_s = 500$ per unit, $\mu = 1$ and $\mu_s = 3$, the transfer function is plotted in Fig. 6 [curve (*b*)]. It follows that the gain has to be increased considerably more than before to obtain stability, and the stability range becomes

$$10\cdot7 < \mu < 50$$

For one and the same load condition, which has been 2·1 per-unit power at unity power factor so far, different values of μ_s lead to different stability ranges, as can be shown in a stability contour diagram (Fig. 7). There the area inside the curve for 2·1 per-unit current represents the stable region for different values of μ and μ_s when the machine operates at 2·1 per-unit power at unity power factor. Similar curves arise for different currents as shown in the same Figure, and it follows that the operating range for an alternator with a high gain in its regulator can be improved by using higher values for μ_s in the stabilizer. For low values of μ the machine tends to become more oscillatory or even unstable with an increase in μ_s.

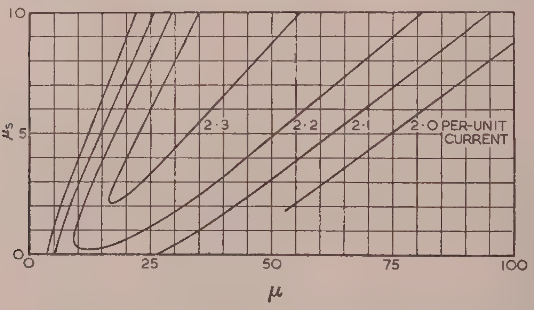


Fig. 7.—Voltage-regulator stability contour diagram for unity power factor, $\tau_s = \tau_e = 500$, $\tau_a = \mu_c = 0$.

Optimum values for μ and μ_s can be chosen from the contours, and $\mu = 20$, $\mu_s = 4$ is a typical set of suitable values. It is seen also that the stability limit can be extended even beyond 2·3 per-unit power by increasing μ and choosing a proper value of μ_s; but the values of μ_s then become too large for practical purposes.

In general, a regulator should provide, for normal loads, a fast non-oscillatory response with small steady-state deviation from the reference value. This requires a high gain μ and a suitable value of μ_s. The contour diagram shows the limitations in gain as regards stability, and a faster response usually means a lower stability limit. The difficulty is to obtain a good performance not only at unity-power-factor loads as considered so far, but also at different power factors. However, conditions satisfying unity power factor in general apply to other power factors as well. This can be seen when comparing Fig. 8 with Fig. 7. In Fig. 8 the stability contours are plotted for 0·7 power factor, and the most suitable sets of values of μ and μ_s correspond to those obtained for unity power factor.

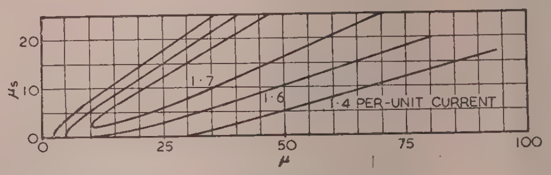


Fig. 8.—Voltage-regulator stability contour diagram for 0·7 power factor; $\tau_s = \tau_e = 500$, $\tau_a = \mu_c = 0$.

(3.2) Regulating Amplifier

So far the time-constant τ_a of the regulating amplifier has been assumed to be zero. In general, the time delay involved is small compared with that in other parts of the circuit, since reliable amplifiers with fast response times are available. Usually a rotating amplifier such as an amplidyne can have a time-constant of about 0·1 sec, and magnetic and electronic amplifiers can provide even faster responses.

As shown in Fig. 6(*a*), the 30 MW alternator when operating near the dynamic limit can be affected seriously if the time-constant τ_a is about 0·5 sec (157 per unit). The machine is unstable, and it is impossible to stabilize it by changing μ. However, for a somewhat smaller time-constant, e.g. 0·1 sec (31·4 per unit), the effect on the transfer function is negligible and only the upper limit in the stability range is reduced somewhat, as shown in curve (*c*). Thus for values of τ_a of about 0·1 sec the effect of the time delay in the amplifier can be neglected in practice.

In general, it is not possible to describe the amplifier performance exactly with a single time delay. However, since its effect is negligible in modern fast continuously acting regulators there is no need to go into greater detail.

(3.3) Controller

By operating on the error the performance of the machine can be improved. For improving the dynamic stability it is usually necessary to introduce some phase-advancing device, e.g. derivative control.

First-derivative error control can be approximated with a derivative transformer giving the following transfer function (see Appendix 9.2):

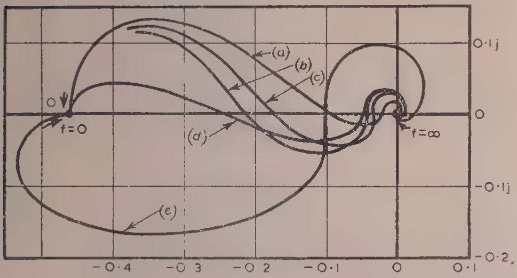
$$F_c = \left(1 + \frac{\mu_c \tau_c}{1 + \tau_c p}\right) \quad . \quad . \quad . \quad . \quad (12$$

μ_c = Transformer ratio

τ_c = Time-constant.

This is the most common control function used for regulator and is taken as a typical example to show the general effect of actual derivative control.

The effect of a derivative controller, with several values of μ



g. 9.—Effect of derivative control on the overall transfer function F for $\tau_e = \tau_s = 500$, $\mu = 1$, $\mu_s = 3$, $\tau_a = 0$.

(a) $\mu_c = 0$ (F without additional control).
(b) $\mu_c = 1$, $\tau_c = 500$.
(c) $\mu_c = 1$, $\tau_c = 250$.
(d) $\mu_c = 1$, $\tau_c = 2\,000$.
(e) $\mu_c = 4$, $\tau_c = 500$.

ad τ_c, on the 30 MW machine can be seen in Fig. 9. The ability improves for low values of the amplifier gain μ, whereas e upper gain limit of the stability range is reduced. The ability ranges are given in Table 1 with the maximum and

Table 1

VARIATION IN CONTROLLER CONSTANTS

τ_c	μ_c	Stability range
500	0	$10{\cdot}7 < \mu < 50$
500	1	$4{\cdot}6 < \mu < 25$
500	4	$2{\cdot}2 < \mu < 10$
250	1	$5{\cdot}7 < \mu < 30$
2 00J	1	$4{\cdot}3 < \mu < 24$

inimum gain specifying this range, and it can be seen that the tio of maximum to minimum gain remains practically unaffected ' changes in μ_c or τ_c. Thus this particular derivative controller es not effectively improve the stability, although it is possible select values of μ_c and τ_c for the most suitable gain obtainable om the amplifier.

(3.4) Power-Angle Regulation

Regulators may be used to control the power angle, in which se the field voltage becomes a function of the power-angle riation $\Delta\delta$. The machine transfer function follows from ns. (1) and (2) by eliminating Δv_t and putting $\Delta T_m = 0$:

$$F_{M\delta} = \frac{\Delta\delta}{\Delta v_{fd}} = \frac{-A_3}{(\tau'_{dz}p + 1)(Mp^2 + Dp + A_1) - A_2} \quad . \quad (13)$$

The regulator transfer function remains the same as for ltage control unless special features are introduced. The gain the amplifier is denoted by a for angle regulation.

With the same regulator as for voltage regulation a stability ntour diagram for variations in a and μ_s is given in Fig. 10 the 30 MW alternator, if it is assumed that $\tau_s = \tau_e = 500$, $= \mu_c = 0$. The diagram is similar to those obtained for ltage control. Although the magnitude of the stability ranges a (for any particular value of μ_s) is smaller than that for μ, actual ratio of the upper to the lower stability limits a_{max}/a_{min} roughly the same numerically as for μ. In the example given, machine with angle regulation is somewhat less stable for same values of μ_s than the voltage-regulated machine, and is generally possible to achieve higher stability with voltage ulation under similar conditions.

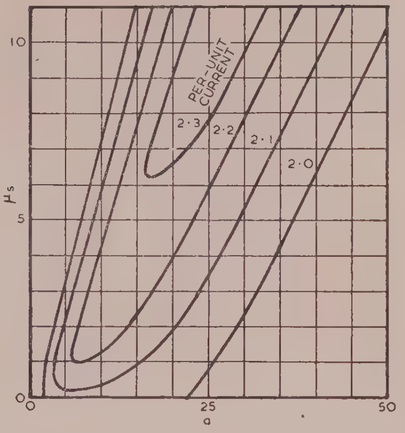


Fig. 10.—Power-angle-regulator stability contour diagram for unity power factor; $\tau_e = \tau_s = 500$, $\tau_a = \mu_c = 0$.

Control proportional to the time derivative of the angle variations can be applied. The results are the same as those obtained for voltage regulation, in that they improve the stability for low feedback gain, but not for high gains.

(4) TRANSIENT ANALYSIS OF SYSTEM EQUATIONS

The transfer-function approach is very useful when determining the stability limits of an alternator. However, when carrying out a detailed investigation a faster method is required. In addition, the transient behaviour of the alternator after a small disturbance has to be determined by solving the complete set of system equations. A complete analysis must therefore be carried out on an automatic computer, and the most suitable is a differential analyser.

As shown in Figs. 14A and 14B, the machine and regulator equations can be set up separately on a differential analyser, and the variations in all the variables can be recorded as required. It is a simple matter to vary the magnitudes of the system parameters such as inertia constants or time-constants and then determine the effects of such alterations. Different kinds of regulators can be tested by changing the regulator set-up without altering the machine set-up. The system may be changed over from voltage to angle regulation by connecting the input of the regulator to the busbar representing the power angle, with a suitable scale-factor modification if necessary.

The set-up for the machine remains unaltered whatever machine is to be simulated. Variations in alternator size are introduced by changes in M, τ'_{dz}, and D, and different load conditions and reactances are introduced by variations in the A_i's, which are all represented by adjustable ratio units.

The machine is set up according to eqns. (1) and (2), which are in a linearized form, and any transients resulting from a disturbance will show how the machine would respond if this disturbance were infinitesimally small. The equations can be assumed as linear only in the close neighbourhood of the steady-state conditions for which the particular A_i's apply. Thus for the response to actual finite disturbances a different alternator set-up would be required, although in general the equations given above can be used to provide approximate results for small finite variations.

(4.1) Alternator Response to Change of Reference Voltage

As a typical example, the 30 MW machine with voltage regulator has been loaded up to 0·8 per-unit current at

0·97 power factor. A sudden change in the reference quantity $\Delta v'_{t0} = +0 \cdot 1$ per unit takes place initially, and the subsequent variations in $\Delta\delta$, Δv_t and Δv_{fd} are recorded for various values of the feedback gain μ and the stabilizing transformer ratio μ_s. The results given in Fig. 11 show the improvement in response for $\mu = 10$ when μ_s is increased from 0 to 3. For $\mu_s = 0$ marked oscillations persist for a period of more than 15 sec, whereas $\mu_s = 3$ leads to a state which is close to critical damping.

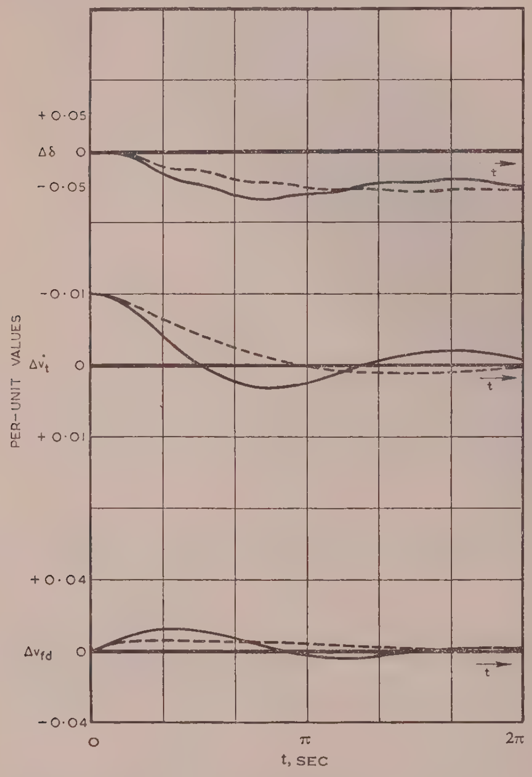


Fig. 11.—Response at normal load.

Current = 0·8 per unit. Power factor = 0·97 (lagging).
$\mu = 40$, $\tau_e = \tau_s = 125$ rad.
----- $\mu_s = 3$
——— $\mu_s = 0$

(4.2) Stability Limit

For determining the stability limit at a particular power factor the load is increased in steps, and for each load the transient response is recorded. The machine is unstable when the magnitude of any one of the oscillations in the system does not decrease with time; thus the limit can be found if a sufficient number of solutions is carried out. This has to be done for several power factors if the complete operating characteristic is required.

Typical machine transients are given in Fig. 12, where the response to a sudden change in $\Delta v'_{t0}$ is recorded for several values of the feedback gain μ for the voltage-regulated 30 MW alternator. An increase in the gain μ reduces the period of the transients, and for low and high values of μ the system is obviously unstable, whereas there is a stable range for medium gains as expected from the stability contour diagrams (see Figs. 7, 8, and 10).

The oscillations arising with a sudden change in $\Delta v'_{t0}$ are due to the hunting effect of the regulator and may be called the feedback oscillations. Their period ranges in general from 1 sec for high feedback gains to 10 sec or more for low gains. The long

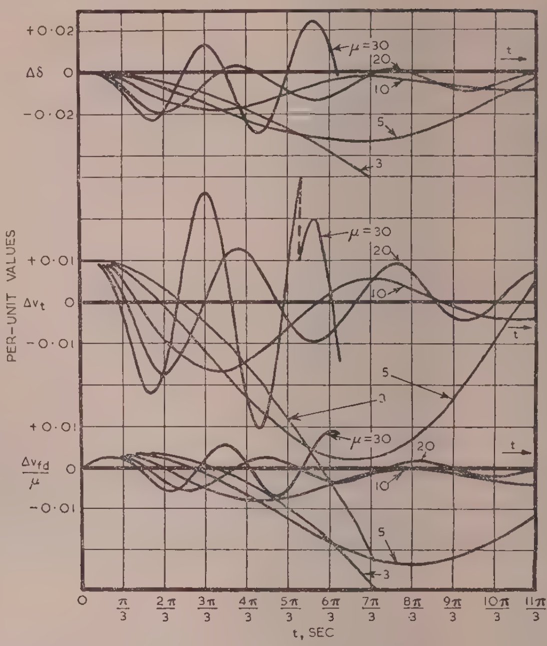


Fig. 12.—Change in response due to different values of μ.

Current = 2·2 per unit. Power factor = 1·0; $\mu_s = 2$

period is possible because of the long time-constant in the fiel circuit.

Another mode of vibration with a relatively short period of th order of 1·5 sec is superimposed on the feedback oscillations o the power angle $\Delta\delta$ in Fig. 11. These oscillations correspon to the natural frequency of the machine without a regulator a determined by the inertia and the steady-state restoring torque Actually, by the use of an error in $\Delta v'_{t0}$, the amplitude of thes natural vibrations has been minimized intentionally, since the would obscure the effects of the regulator.

From Fig. 12 the general conclusion follows that instabilit for low feedback gains is due to the slow response of the regulato Thus the stability could be improved by reducing the time dela in the feedback path. However, this would mean a reduction i field time-constant, which is generally fixed for a particul machine size.

When increasing the gain the period of the feedback oscill tions decreases, as shown in Fig. 12, until it approaches that the natural frequency, and the resulting resonance leads t instability. Thus the high gain stability limit is determined b an interaction between the feedback oscillation and the natur vibrations of the machine considered. Following from this, th high gain stability limit could be extended by increasing th period of hunting. This can be achieved by increasing th alternator-field time-constant. A reduction in the period of th natural frequency would have the same effect and could b obtained by reducing the rotational inertia of the machine.

(5) CONTROL OF PRIME-MOVER TORQUE

The torque of a prime mover changes with speed and so affec the performance of the alternator. In addition, when allowan is made for governor action, the variations in the mechanic torque ΔT_m can be expected to modify the stability limits.

The prime-mover output-torque variation is given as

$$\Delta T_m = \frac{\partial T_m}{\partial \omega}\Delta\omega = \frac{\partial T_m}{\partial \omega}p\Delta\delta \quad . \quad . \quad . \quad . \quad (1$$

here ω = Instantaneous angular speed $= \omega_0 + \Delta\omega$

ω_0 = Synchronous speed $= 2\pi f_0$

$\Delta\omega$ = Small change in ω

$\dfrac{\partial T_m}{\partial \omega}$ = About -1 (see Reference 12).

This effect is allowed for in the general machine equation (1) ı the damping factor D. In eqn. (14) the value of D, which 'as taken as 3, had little effect on the solutions, and it follows ıat the damping effect of the prime mover itself is negligible.

(5.1) Speed Governor

By introducing a speed governor and neglecting the time delays ı the governing system, the mechanical torque variation ecomes

$$\Delta T_m = -\frac{1}{r}p\Delta\delta + \frac{\partial T_m}{\partial\omega}p\Delta\delta \quad . \quad . \quad . \quad (15)$$

$$\simeq -\left(\frac{1}{r}+1\right)p\Delta\delta$$

'here r = Regulation of governor.

Thus the torque relation for the machine as expressed by qn. (1) changes into

$$(\tau'_{dz}p + 1)\left[Mp^2 + \left(D + \frac{1}{r} + 1\right)p + A_1\right] - A_2\Big\}$$

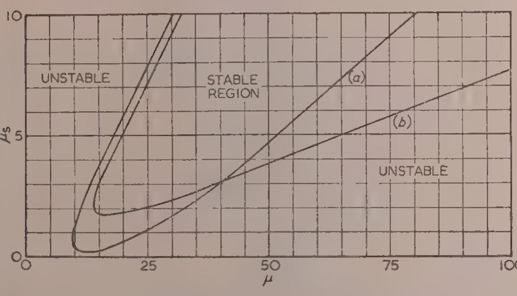
$$\Delta\delta + A_3\Delta v_{fd} = 0 \quad . \quad (16)$$

vith an effective increase in the damping factor D to

$$D' = \left(D + \frac{1}{r} + 1\right) \quad . \quad . \quad . \quad . \quad . \quad (17)$$

This change will cause a more effective damping of any oscillaions of the machine. However, it does not affect the stability ondition [eqn. (4)] as obtained for a machine with disconinuous regulation.

For a regulated machine, an increase in D can cause a coniderable change in the dynamic stability limit. In Fig. 13 the ffect of a governor with 5% regulation is shown for the voltageegulated 30 MW alternator. As can be seen, the stability decreases for the normal values of the stabilizer constant μ_s, whereas it improves for large values of μ_s.



ig. 13.—Effect of speed governor on the stability for unity power factor and 2·2 per-unit current considering several values of μ and μ_s ($\tau_s = \tau_e = 500$, $\tau_a = \mu_c = 0$, $\tau_1 = \tau_2 = 0$, $R = 5\%$).

(a) Voltage regulator only.
(b) Voltage regulator and governor.

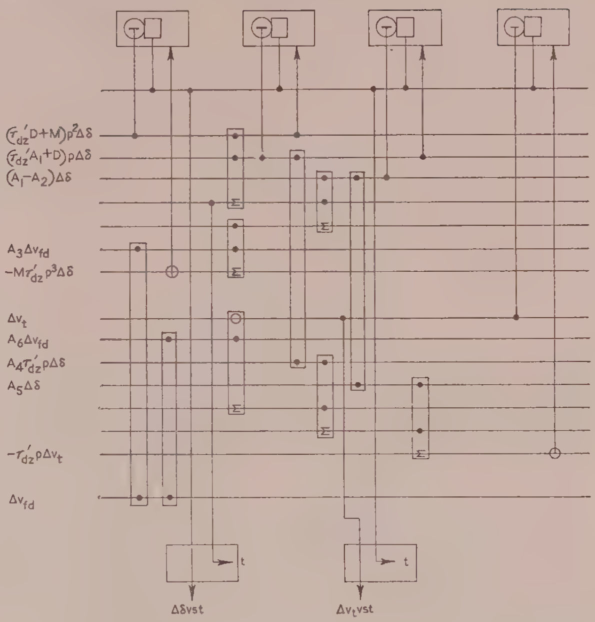


Fig. 14A.—Differential analyser schematic for the alternator.

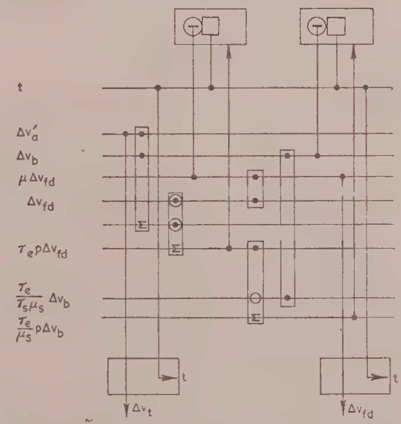


Fig. 14B.—Differential analyser schematic for the regulator.

(5.2) Tie-Line Power Control

When the prime-mover torque is controlled with the tie-line power, and time delays in the controller are neglected, it is possible to improve the damping of system oscillations.

The electrical torque appears in eqn. (1) as

$$\Delta T_{e1} = \left[A_1 - A_2\left(\frac{1}{\tau'_{dz}p+1}\right)\right]\Delta\delta - \frac{A_3}{\tau'_{dz}p+1}\Delta v_{fd} \quad . \quad (18)$$

Neglecting time delays the mechanical torque becomes

$$\Delta T_m = \frac{\Delta T_{e1}}{r} = \frac{1}{r}\left[A_1 - \left(\frac{A_2}{\tau'_{dz}p+1}\right)\right]\Delta\delta - \frac{1}{r}\frac{A_3}{\tau'_{dz}p+1}\Delta v_{fd}$$

Thus the torque equation follows as

$$\left\{(\tau'_{dz}p + 1)\left[Mp^2 + Dp + A_1\left(1 + \frac{1}{R}\right)\right] - A_2\left(1 + \frac{1}{r}\right)\right\}\Delta\delta - A_3\left(1 + \frac{1}{r}\right)\Delta v_{fd} = 0 \quad . \quad (19)$$

This means that A_1 and A_2 change into the effective constants

$$A'_1 = A_1\left(1 + \frac{1}{r}\right) \quad . \quad . \quad . \quad . \quad . \quad (20)$$

$$A'_2 = A_2\left(1 + \frac{1}{r}\right)$$

From eqn. (1) it can be seen that this is leading to the same stability condition as for the machine alone [see eqn. (4)], since the regulation r is positive. Thus the steady-state stability limit remains unaffected by tie-line power control if the time delays are neglected.

(5.3) Time Delay in Governing System

The time-constants in governing systems are comparable with those in the regulator. They are not negligible in general and lie within the range [12,13] 0·5–10 sec. The simplest description of a governor requires at least a second-order delay[13] and is of the form

$$\Delta T_m = \frac{-1}{(1 + \tau_1 p)(1 + \tau_2 p)}\left(\frac{p\Delta\delta}{r} + \frac{G}{r}\right) \quad . \quad . \quad (21)$$

where τ_1, τ_2 = Governor time-constants

$$G = F_{TL} \times \Delta T_{e1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

F_{TL} = Delay function for tie-line power controller.

It follows that the characteristic equations for the machine system will increase by an order of two in the simplest case. This makes computations rather complex, and the problem must be handled on an automatic computer.

In general, the time delay reduces the damping effect. However, it also delays the damping, and as the delay time-constant approaches the period of natural vibrations in the system the damping due to the controller becomes positive and makes the system more unstable.

The natural period of the 30 MW alternator used as an example so far was about 1 sec. Thus a delay time-constant of about 1 sec in the governor introduces positive damping, and even for small values of r the alternator can become unstable under normal operating conditions. A regulation of the order of 5% can be critical.

(6) CONCLUSIONS

The overall steady-state or dynamic stability range is to some extent inversely proportional to the synchronous reactance of the alternator, as can be deduced from eqn. (5) and Fig. 4. This applies particularly to the condition when the external reactance X_e is small. Thus greater stability without continuous regulation can be achieved only by reducing the synchrous reactance, which involves an increase in frame size of the machine, with a considerable increase in cost.

When using a continuously-acting regulator, expensive alternators with large frames can be avoided, which reduces the overall cost. In general, the stability range of a machine can be increased by 50–100% with continuous voltage regulation, even for power factors as low as 0·3.

The actual change in the dynamic limit depends on the various components in the controlling circuits, and the choice of the circuit parameters is important. The gain in the feedback loo is the most critical factor, with an optimum value for any set c parameters. Thus, once the gain is increased beyond a definit value, the dynamic stability limit will drop, although the stead state error continues to decrease. The dynamic limit for hig feedback gains can be improved by an increase in the time dela in the feedback loop or by a decrease in the alternator inertia.

In general it has been found also that the choice of the para meters in the feedback loop is practically independent of powe factor.

Speed governing increases the effective damping of an alternato if time delay is neglected. In that case it will not affect th steady-state stability limit, although it does modify the dynam limit.

Tie-line power control also increases the effective damping an reduces the period of natural vibrations of an alternator. If tim delays are neglected the controller affects only the dynamic limi

When allowance is made for time delay in the governing systen a detailed investigation into the modification of the dynam stability limit becomes rather tedious. For a speed governo the time delays are usually of such a magnitude as to introduc positive damping and reduce the stability limit. Tie-line powe controllers can be used for stabilizing, but a detailed analys is always necessary to determine their actual performance.

(7) ACKNOWLEDGMENTS

The authors would like to thank Prof. D. M. Myers for the us of the differential analyser of the C.S.I.R.O. Mathematic Instruments Section, Sydney, and Mr. J. Foster of that Sectio for his assistance in the operation of the differential analyse They also wish to express their appreciation to Mr. J. T. Roll for his interest and constructive comments when reading th manuscript.

The work described in the paper resulted from a joint investiga tion carried out at the University of Sydney for the Electrici Commission of New South Wales.

(8) REFERENCES

(1) WALKER, J. H.: "Operating Characteristics of Salient-Po Machines," *Proceedings I.E.E.*, Paper No. 1211 February, 1953 (**100**, Part II, p. 13).

(2) CONCORDIA, C.: "Steady State Stability of Synchrono Machines as affected by Angle-Regulator Chara teristics," *Transactions of the American I.E.E.*, 1948, **6** Part I, p. 687.

(3) CONCORDIA, C.: "Steady State Stability of Synchrono Machines as affected by Voltage-Regulator Chara teristics," *ibid.*, 1944, **63**, p. 215.

(4) CONCORDIA, C.: "Effect of Buck-Boost Voltage Regulat on Steady State Stability Limit," *ibid.*, 1950, **69**, Part p. 380.

(5) HEFFRON, W. G., and PHILLIPS, R. A.: "Effect of Mode Amplidyne Voltage Regulators on Underexcited Oper tion of Large Turbine Generators," *ibid.*, 1952, **71**, p. 6

(6) FARNHAM, S. B., and SWARTHOUT, R. W.: "Field Excitati in Relation to Machine and System Operation," *Pow Apparatus and Systems*, 1953, No. 9, p. 1215.

(7) ADKINS, B.: "Analysis of Hunting by Means of Vect Diagrams," *Journal I.E.E.*, 1946, **93**, Part II, p. 541.

(8) PARK, R. H.: "Two-Reaction Theory of Synchronc Machines, Generalized Method of Analysis—Part *Transactions of the American I.E.E.*, 1929, **48**, p. 716.

(9) THALER, G. J., and BROWN, R. G.: "Servomechani Analysis" (McGraw-Hill, 1953).

(10) Myers, D. M., and Blunden, W. R.: "The C.S.I.R.O. Differential Analyser," *Journal of the Institution of Engineers, Australia*, Oct.–Nov., 1952, **24**, p. 3.

(11) Crary, S. B.: "Power System Stability, Vol. II" (John Wiley & Sons, New York, 1945).

(12) Ruedenberg, R.: "The Frequency of Natural Power Oscillations in Interconnected Generating and Distribution Systems," *Transactions of the American I.E.E.*, 1943, **62**, p. 791.

(13) Concordia, C., and Kirchmayer, L. K.: "Tie-Line Power and Frequency Control of Electric Power Systems," *Power Apparatus and Systems*, 1953, No. 6, p. 562.

(9) APPENDICES

(9.1) Derivation of Machine Equations

The general voltage equations for the machine without damper windings as shown in Fig. 15 (see also References 3 and 6) are:

$$\begin{bmatrix} v \\ v_d \\ v_q \end{bmatrix} = \begin{bmatrix} R_f + X_f p & -X_{afd}p & 0 \\ X_{afd}p & -X_d p & X_q p\theta \\ X_{afd}p & -X_d p\theta & -X_q p \end{bmatrix} \times \begin{bmatrix} i_f \\ i_d \\ i_q \end{bmatrix}$$

$$= \begin{bmatrix} R_f\dfrac{i_f}{\Phi_{ffd}} + p & 0 & 0 \\ 0 & p & -p\theta \\ 0 & p\theta & p \end{bmatrix} \times \begin{bmatrix} \Phi_{ffd} \\ \Phi_d \\ \Phi_q \end{bmatrix} \qquad (23)$$

neglecting armature resistance,

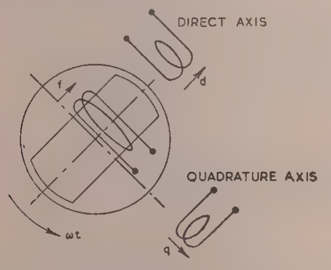


Fig. 15.—Direct and quadrature representation of a synchronous machine.

and

$$\begin{bmatrix} \Phi_{ffd} \\ \Phi_d \\ \Phi_q \end{bmatrix} = \begin{bmatrix} X_f & -X_{afd} & 0 \\ X_{afd} & -X_d & 0 \\ 0 & 0 & -X_q \end{bmatrix} \times \begin{bmatrix} i_f \\ i_d \\ i_q \end{bmatrix} \quad . \quad . \quad (24)$$

It is usually more convenient to express the field voltage v in terms of the per-unit field voltage v_{fd} required to produce a per-unit flux linking the field and the armature.

Thus
$$v_{fd} = \frac{X_{afd}}{R_f}v = X_{afd}i_f + \frac{X_{afd}}{R_f}p\Phi_{ffd}$$
$$= X_{afd}i_f + \tau'_{d0}p\left(\frac{X_{afd}}{X_f}\Phi_{ffd}\right)$$
$$= i_{fd} + \tau'_{d0}p\Phi_{fd} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

where i_{fd} is the field current referred to the armature and $\Phi_{fd} = \dfrac{X_{afd}}{X_f}\Phi_{ffd}$

From eqn. (24)

$$\Phi_{ffd} = \frac{X_f}{X_{afd}}\left[\Phi_d + \left(X_d - \frac{X^2_{afd}}{X_f}\right)i_d\right] = X_f i_f - \frac{X_f}{X_{afd}}(X_d - X'_d)i_d$$

where
$$X'_d = X_d - \frac{X^2_{afd}}{X_f}$$

Thus
$$\Phi_{fd} = \frac{X_{afd}}{X_f}\Phi_{ffd} = i_{fd} - (X_d - X'_d)i_d \quad . \quad . \quad (26)$$

If only small variations in the variables are considered, the following equations arise:

$$\begin{bmatrix} \Delta v_{fd} \\ \Delta v_d \\ \Delta v_q \end{bmatrix} = \begin{bmatrix} \dfrac{\Delta i_{fd}}{\Delta\Phi_{fd}} + \tau'_{d0}p & 0 & 0 \\ 0 & p & -p\theta \\ 0 & p\theta & p \end{bmatrix} \times \begin{bmatrix} \Delta\Phi_{fd} \\ \Delta\Phi_d \\ \Delta\Phi_q \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\Delta i_{fd}}{\Delta\Phi_{fd}} + \tau'_{d0}p & 0 & 0 \\ 0 & p & -1-p\Delta\delta \\ 0 & 1 + p\Delta\delta & p \end{bmatrix} \begin{bmatrix} \Delta\Phi_{fd} \\ \Delta\Phi_d \\ \Delta\Phi_q \end{bmatrix} \qquad (27)$$

Thus approximately, by neglecting the variations in the derivatives,

$$\begin{bmatrix} \Delta v_{fd} \\ \Delta v_d \\ \Delta v_q \end{bmatrix} = \begin{bmatrix} \dfrac{\Delta i_{fd}}{\Delta\Phi_{fd}} + \tau'_{d0}p & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \Delta\Phi_{fd} \\ \Delta\Phi_d \\ \Delta\Phi_q \end{bmatrix} \quad . \quad (28)$$

Also

$$\begin{bmatrix} \Delta\Phi_{fd} \\ \Delta\Phi_d \\ \Phi\Delta_q \end{bmatrix} = \begin{bmatrix} 1 & -(X_d - X'_d) & 0 \\ 1 & -X_d & 0 \\ 0 & 0 & -X_q \end{bmatrix} \times \begin{bmatrix} \Delta i_{fd} \\ \Delta i_d \\ \Delta i_q \end{bmatrix} \quad . \quad (29)$$

For the terminal voltage v_t with

$$v_t^2 = v_d^2 + v_q^2$$

we get, using Fig. 16,

$$\Delta v_t = \frac{v_{d0}}{v_{t0}}\Delta v_d + \frac{v_{q0}}{v_{t0}}\Delta v_q \quad . \quad . \quad . \quad . \quad (30)$$

For the torque the following expression arises

$$[Mp^2 + (Dp - 1)]\theta = T_m - T_{el} = T_m - \Phi_d i_q + \Phi_q i_d$$

or in terms of small variations

$$(Mp^2 + Dp)\Delta\delta = \Delta T_m$$
$$- (\Phi_{d0}\Delta i_q + i_{q0}\Delta\Phi_d - \Phi_{q0}\Delta i_d - i_{d0}\,\Delta\Phi_q) \quad . \quad (31)$$

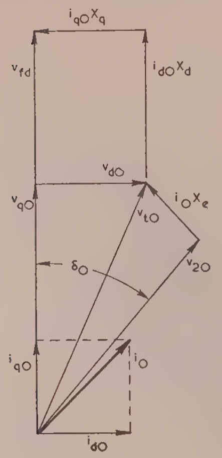


Fig. 16.—Steady-state vector diagram for synchronous machine.

For an external reactance X_e directly connected to an infinite busbar, two more equations are given when assuming $\Delta v_2 = 0$:

$$\left.\begin{aligned}\Delta v_d &= -X_e\Delta i_q + (v_{20}\cos\delta_0)\Delta\delta\\ \Delta v_q &= X_e\Delta i_d - (v_{20}\sin\delta_0)\Delta\delta\end{aligned}\right\} \quad . \quad . \quad (32)$$

Thus there are ten equations [eqns. (28)–(32)] with 12 unknowns, two of which (i.e. ΔT_m and Δv_{fd}) are independent. The equations can be solved, therefore, if the boundary conditions are known. For convenience, all variables except Δv_t, $\Delta\delta$, Δv_{fd} and ΔT_m are eliminated by means of the steady-state vector diagram (Fig. 16), and this leads to the following two equations:

$$\left.\begin{aligned}&[(\tau'_{dz}p + 1)(Mp^2 + Dp + A_1) - A_2]\Delta\delta\\ &\qquad\qquad + A_3\Delta v_{fd} = (\tau'_{dz}p + 1)\Delta T_m\\ &(A_4\tau'_{dz}p + A_5)\Delta\delta + A_6\Delta v_{fd} = (\tau'_{dz}p + 1)\Delta v_t\end{aligned}\right\} \quad . \quad (33)$$

with

$$\left.\begin{aligned}A_1 &= \frac{v_{20}v_{q0}\cos\delta_0}{X_e + X_q} + \frac{v_{20}v_{d0}\sin\delta_0}{X_e + X'_d}\left(1 - \frac{X'_d}{X_q}\right)\\ A_2 &= \frac{v_{20}^2\sin^2\delta_0(X_d - X'_d)}{(X_e + X'_d)(X_e + X_d)}\\ A_3 &= \frac{v_{20}\sin\delta_0}{(X_e + X_d)}\\ A_4 &= \frac{v_{d0}}{v_{t0}}v_{20}(\cos\delta_0)\frac{X_q}{(X_e + X_q)}\\ &\qquad - \frac{v_{q0}}{v_{t0}}v_{20}(\sin\delta_0)\frac{X'_d}{(X_d + X_e)}\\ A_5 &= -\frac{v_{20}v_{q0}}{v_{t0}}(\sin\delta_0)\frac{X_d}{(X_e + X_d)}\\ &\qquad + \frac{v_{20}v_{d0}}{v_{t0}}(\cos\delta_0)\frac{X_q}{(X_e + X_q)}\\ A_6 &= \frac{v_{q0}}{v_{t0}} \times \frac{X_e}{(X_e + X_d)}\end{aligned}\right\} \quad . \quad (34)$$

(9.2) Derivation of Regulator Equations

The open-loop transfer function F_R of the regulator as shown in Fig. 3 is made up of three components:

$$F_R = F_cF_aF_r$$

The transfer function of the controller, F_c, depends on the method of control employed. In Fig. 3 a derivative control is applied by means of a derivative transformer for which the following two equations apply for the primary and secondary voltages:

$$\left.\begin{aligned}\Delta v_p &= (R_p + L_pp)i_p + Mpi_c\\ \Delta v_c &= Mpi_p + (R_c + M_cp)i_c\end{aligned}\right\} \quad . \quad . \quad . \quad (35)$$

Solving for Δv_c and putting

$$M^2 = L_pL_c$$

$$\frac{M}{R_p} = \sqrt{\left(\frac{L_c}{L_p}\right)} \times \frac{L_p}{R_p} = \frac{N_c}{N_p}\frac{L_p}{R_p}$$

where N_p = Number of primary turns, N_c = Number of secondary turns, and

$$\frac{N_c}{N_p} = \mu_c; \quad \frac{L_p}{R_p} = \tau_c$$

If i_c is small

$$\Delta v_c = \frac{\mu_c\tau_cp}{1 + \tau_cp}\Delta v'_t \quad . \quad . \quad . \quad . \quad . \quad (36)$$

The controller transfer function then becomes

$$F_c = \left[1 + \frac{\mu_c\tau_cp}{(1 + \tau_cp)}\right] \quad . \quad . \quad . \quad . \quad (37)$$

For the amplifier a single delay is assumed with a time-constant τ_a, giving

$$F_a = \frac{\mu}{(1 + \tau_ap)} \quad . \quad . \quad . \quad . \quad . \quad (38)$$

For the exciter circuit the following two equations apply:

$$\Delta v_b = \frac{\mu_s\tau_sp}{1 + \tau_sp}\Delta v_{fd}$$

$$\Delta v_{fd} = \frac{-1}{1 + \tau_ep}(\Delta v'_a + \Delta v_b) \quad . \quad . \quad . \quad (39)$$

Thus $F_r = \dfrac{\Delta v_{fd}}{\Delta v'_a} = \dfrac{-(1 + \tau_sp)}{\{1 + [\tau_e + \tau_s(\mu_s + 1)]p + \tau_e\tau_sp^2\}}$. (40)

and

$$F_R = \left[\frac{-\mu}{(1 + \tau_ap)}\right]\left(1 + \frac{\mu_c\tau_cp}{1 + \tau_cp}\right)$$

$$\left\{\frac{(1 + \tau_sp)}{1 + [\tau_e + \tau_s(\mu_s + 1)]p + \tau_e\tau_sp^2}\right\} \quad . \quad (41)$$

[A discussion on the above paper will be found on page 231.]

621.315.615.2 : 621.3.015.5

Monograph No. 135 S
June 1955

THE ELECTRIC STRENGTH OF TRANSFORMER OIL

By M. E. ZEIN EL-DINE, Ph.D., and H. TROPPER, Ph.D., Associate Member.

(*The paper was first received* 30*th June*, 1954, *and in revised form* 4*th February*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in June*, 1955.)

SUMMARY

The paper gives results of electric breakdown tests on transformer oil, made with direct and alternating voltages and with impulses of different durations. The dried oil used in these tests was filtered and carefully degassed in a closed test apparatus which incorporated the test cell. The oil treated in this way was tested with uniform and non-uniform electrode configurations, and a number of factors which influence the electric strength of the treated oil were examined.

With the electric strength of the treated oil as a reference, the effect on its value was examined when a number of different impurities were added to the treated oil.

Finally, a theory of the electric breakdown of this type of liquid is suggested which conforms to the experimental observations reasonably well.

(1) INTRODUCTION

The liquid dielectrics used for insulating purposes in industry, such as transformer oil, are of a composite nature and always contain impurities in various amounts. They readily absorb moisture and gases, and the impurities, although normally small in amount, nevertheless lower the electric strength and cause considerable scatter of the breakdown measurements. This makes it difficult to correlate the experimental results of different investigators.

Such a correlation would be greatly facilitated if the impurity content of the liquid that is being tested could be determined. This, however, is difficult and, in the absence of any practicable method, was obviated by adopting a particular test procedure in the experiments. This consisted in subjecting the oil to a number of cleaning operations in order to obtain as high an electric strength as possible; the electric strength of this "treated" oil was then used as a criterion for the state of purity, and in this way the effect of adding impurities could be conveniently studied. The standard purification technique used throughout the experiments consisted in a prolonged drying of the oil in the presence of metallic sodium wire and silica gel. The oil was then degassed and filtered in a closed all-glass system of which the test cell formed an integral part. It was found that the oil treated in this way had a high electric strength which was very sensitive to changes in the impurity content: the addition of even small amounts of impurities resulted in an appreciable lowering of the breakdown values.

The tests on the treated oil were made with direct and alternating voltages and, to eliminate time-dependent breakdown mechanisms, voltage impulses with durations of the order of microseconds were also used. A number of factors were examined which are known to affect the measured electric strength values. These include the dependence of the electric strength on the metal of the test electrodes and on their size both for uniform and non-uniform electric fields. In non-uniform fields the polarity effect was examined and it was shown how it was influenced by the addition of various kinds of impurities.

Several different explanations for the electric breakdown in liquid dielectrics have been put forward from time to time. Most of these suggest in one way or another that the breakdown takes place in gas bubbles which form either from the gas contained in the liquid or by vaporization of the liquid itself. These explanations assume that the breakdown is of the gaseous type and therefore should depend on the external pressure. To discriminate between these "bubble" mechanisms and others which involve breakdown in the liquid phase, the pressure-dependence of the breakdown voltage was also examined. The experiments were made with both uniform and non-uniform electrode gaps.

(2) APPARATUS AND EXPERIMENTAL PROCEDURE

(2.1) Supplies and Measurement of High Voltages

A description of the voltage sources used for the tests with direct and alternating voltages and the measurement of these voltages has been given in a previous publication[19] and will be omitted here. Reference to that publication should also be made for a description of the precautions taken during the tests to limit the discharge current which passed through the test liquid during breakdown.

For the tests with impulse voltages a single-stage generator was used. The generator capacitance was $0{\cdot}017\,\mu$F and a load capacitance of $300\,\mu\mu$F was connected in parallel with the test cell to swamp the effect of any capacitance changes in the test load. With this capacitance a constant wavefront of approximately 1 microsec was obtained. The tail of the impulses was varied by changing the value of the tail resistance of the circuit. Impulses having tails of 3, 5, 10, 25, 75 and 100 microsec were used during the experiments. The waveshapes were recorded by a cathode-ray oscillograph, and this instrument was also used to determine the crest value of the impulses. To simplify the test procedure the voltage across the generator capacitance was calibrated in terms of the crest value of the impulses. The difference between these two voltages was small; even in the worst case, for 1/3 microsec impulses, it was only 7%.

(2.2) Test Apparatus

An all-glass test apparatus was constructed for carrying out the following operations on the oil before subjecting it to the breakdown tests:

(*a*) Degassing of the oil.
(*b*) Filtering of the oil.
(*c*) Filling of the test cell under vacuum in order to avoid contamination of the treated oil.

Fig. 1 shows the general layout of the apparatus. It consists of three parts, namely a circulating degassing system, a filtration system and the test cell.

In the course of the work a number of sintered glass filters of different porosities were used; the porosities of these filters and the corresponding grades were as follows:

Grade	1	2	3	4
Porosity, microns . .	90–100	40–90	15–40	5–15

Correspondence on Monographs is invited for consideration with a view to publication.

The paper is based on a thesis submitted for the degree of Doctor of Philosophy at London University.

Dr. Zein El-Dine is in the Faculty of Engineering, University of Alexandria. Dr. Tropper is in the Electrical Engineering Department, Queen Mary College, University of London.

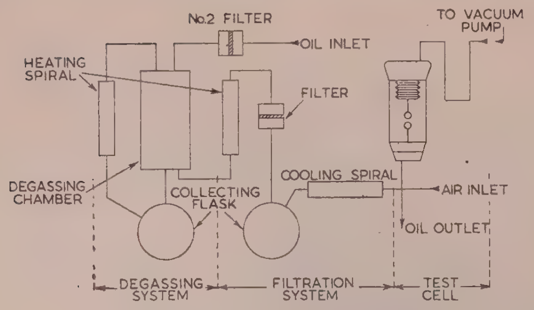


Fig. 1.—Layout of test apparatus.

(2.2.1) Degassing System.

For degassing, the pre-heated oil was introduced under vacuum into the degassing chamber of the apparatus through a sintered glass filter of grade No. 2 and a nozzle of 1 mm diameter. After leaving the nozzle the oil hit a glass baffle which produced an effect similar to that of mechanically shaking the oil. From the baffle plate the oil fell to the bottom of the degassing chamber in a number of fine jets. It could be transferred from there to a collecting flask and reintroduced into the degassing chamber through a heating jacket and another nozzle, and this degassing cycle could be repeated any number of times.

(2.2.2) Filtration System.

For the filtration, the degassed oil was collected at the bottom of the degassing chamber, and the vacuum was then applied to the filter and the remaining parts of the apparatus. A small amount of dried and filtered air was admitted to the degassing chamber, sufficient to provide the pressure necessary to force the oil through the filter. Since the amount of gas absorbed by the liquid depends on its partial pressure, it may be assumed that only very little was re-absorbed at this stage. Moreover, some of this was probably removed again during the filtration.

To facilitate filtration when very fine filters were used, the oil was first heated by passing it through a glass spiral arranged in a heating jacket.

After filtration, the oil was usually left to cool under vacuum in the collecting flask. The test cell was evacuated before filling and the oil was passed into it through a glass spiral, where it could be further cooled if necessary.

The heating of the oil during the various stages of its treatment was by hot water, which was circulated through the heating jackets by means of a small motor-driven pump.

The vacuum-pump connections to the apparatus were through vapour traps which were cooled with solid carbon dioxide, and the pressure was measured by a vacuostat. All ground-glass joints were sealed with a high-vacuum C.A. grease as recommended by Polley.[1] For organic liquids Polley has demonstrated the superiority of this type of grease over the silicone grease available commercially. As a precaution, the grease was applied only to the outer half of the ground surface.

All connections of the apparatus to atmosphere were made through sintered glass filters of grade No. 4 and drying tubes containing phosphorous pentoxide and glass wool.

(2.2.3) Test Cell.

The test cell consisted essentially of a cylindrical glass container having an inside diameter of 50 mm and a length of 120 mm, in which the electrodes were arranged vertically. The lower electrode was carried by a stainless-steel shank of $\frac{5}{16}$ in diameter which was screwed into a chromium-plated steel plate. This formed the upper part of the base. The lower part was of conical shape and tapered so that it fitted accurately into the tapered bottom of the glass container. There was a clearance between the two parts, and by means of three screws it was possible to adjust their relative positions so that the lower electrode could be accurately aligned with the top electrode.

The top electrode shank was not directly connected to the micrometer, but fixed to stainless-steel bellows. In this way the electrode gap could be adjusted with the cell completely sealed off from the atmosphere.

For the evacuation of the cell a glass tube was connected to the top of the glass container. Another glass tube with a 3-way tap and ground sockets was joined to the bottom of it. One of the sockets was used to connect the test cell to the remainder of the test apparatus for filling purposes and the other served as an outlet for the cell. Finally, provision was made to connect the inside of the cell to the atmosphere through a sintered glass filter of grade No. 4 and a drying tube containing phosphorous pentoxide and glass wool.

The cell was always used completely filled with test liquid so that the liquid was visible in the glass tube at the top of the cell. Into this tube, which was normally connected to the vacuum pump, a few drops of re-distilled mercury were introduced when carrying out high-pressure tests. This was done to seal the test liquid from the gas of the pressure vessel.

Normally, the electrode gap was adjusted after the cell was filled with oil. The zero reading of the gap was determined by connecting the electrodes in series with a source of negative grid voltage in the grid circuit of a triode. When the electrodes were brought into contact, the reading of a milliammeter in the anode circuit suddenly dropped to zero. This method is very sensitive and has the advantage that no current passes through the electrodes when they are in contact.

The usual gap settings, which were of the order of 0·25 mm, were measured with the micrometer to an accuracy of about 1%. For very small gaps an optical lever system was used. The sensitivity of this method was estimated to be approximately 1·47 cm/micron, and it was possible to measure gaps as small as 15 microns with an estimated accuracy of 5%. For larger gaps than this the accuracy improved very quickly.

(2.3) Cleaning Technique

Before assembly, all glass parts of the test apparatus were treated with concentrated sulphochromic acid and scrubbed with hot water and a detergent. They were then given a rinse with tap water followed by a rinse with distilled water which had been filtered through a sintered glass filter of grade No. 4. Finally they were dried for three hours in the presence of calcium chloride in an electric oven at a temperature of 110° C.

The metal parts of the cell were washed in trichlorethylene, rinsed with acetone and finally rinsed with filtered acetone. They were then dried for a few minutes in the electric oven.

After assembling the whole test apparatus, including the test cell, the usual procedure was to apply a vacuum of 0·5–1·0 mm Hg for about three hours, during which time the different parts of the system were warmed by blowing hot air on the outer walls. This was done to remove the moisture which was adsorbed during the assembly. As a further precaution the first filling was used for rinsing the test cell and was then discarded. A source of strong light placed behind the cell made it possible to verify that there were no fine fibres in the electrode gap.

Throughout the experiments the electrode surfaces were periodically examined. To obtain smooth surfaces the electrodes were buffed with a soft mop using a suitable grease. To remove the grease they were washed first with trichlorethylene and then in a boiling mixture of ether and acetone. After a careful rinse ı filtered distilled water they were put in the oven for drying, eft there to cool and then immediately assembled.

(2.4) General Test Procedure

The oil used throughout the experiments complied with ..S. 144: 1951. Several weeks before the oil was tested it was lrawn from the barrel and stored in 2·5-litre glass bottles conaining metallic sodium wire. A few days later some selfndicating silica gel, size 20, was added to the oil in each bottle. The bottles were frequently shaken during storage and if the urface of the sodium appeared to be oxidized, or if the colour of the silica gel had changed, further quantities of one or both of these substances were added.

As to the subsequent treatment of the oil in the test apparatus, extensive preliminary tests were made in order to examine the lependence of the electric strength on such factors as the temperature and pressure used in the degassing and filtration processes and the number of degassing cycles. This work showed hat the effectiveness of the degassing depended very much on he temperature of the oil and the pressures used. The lower he pressure and the higher the oil temperature the more effective he degassing. For a given filter grade, on the other hand, the iltration process did not depend to the same extent on the oil emperature and pressure. It was thought advisable not to use for the degassing a vacuum higher than 1 mm Hg and a temperature higher than 90° C so that the partial evaporation of the lighter constituents and oxidation of the oil would not take place. Under these conditions it was found that, even after several hundred degassing cycles, the amount of liquid in the cold trap of the vacuum system was negligible.

Experiments in which the number of degassing cycles was varied showed that, with the method adopted, the increase in electric strength was most marked during the first five to ten cycles. Further cycles did not produce an appreciable increase n strength, and a point was soon reached when the change in electric strength was comparable to the variation of the individual measurements. These tests also revealed an important relation between the gas content of the oil and the conditioning effect, i.e. the tendency of the electric strength to increase with ncreasing numbers of breakdowns. Whereas, after only a few degassing cycles the electric strength of the oil increased continually with subsequent breakdowns, such an increase was found only during a small number of initial breakdowns when he oil was well degassed.

Accordingly, in all the experiments the standard treatment of he oil before the breakdown tests consisted of ten degassing cycles at a pressure of approximately 0·5 mm Hg, with the oil at a temperature of 85–90° C. These temperatures and pressures were chosen also for the filtration, for which a sintered glass filter of grade No. 4 was used. The oil obtained in this way will be referred to as treated oil. Although it contained a certain amount of moisture and gas, as well as other impurities, its purity was fairly high. Throughout the work the state of purity of this treated oil was regarded as a convenient reference by which the effect of adding impurities could be judged. It is believed that only by adopting such a procedure is it possible to ascertain the extent to which various factors influence the breakdown mechanism of a complex dielectric liquid such as transformer oil.

The electric strengths which were obtained with treated oil were very high, and as far as the authors are aware similar high values have hitherto been reported only by Race.[3] In all the tests several a.c. conditioning discharges were passed through the test samples before the breakdown voltages were recorded. For spherical electrodes the electric strengths stated were determined by means of the Russel[4] coefficient.

For impulse tests the experimental procedure was to apply first a voltage of about 60% of the estimated breakdown voltage and to increase it in steps of about 0·5 kV. At every stage in this procedure 3 to 5 applications were made. After breakdown, the test was repeated at least five times at suitable intervals, in order to establish the true breakdown value. The measurements with this type of voltage were very consistent, and reliable results could be obtained from few individual measurements.

(3) EXPERIMENTAL WORK

(3.1) Effect of Duration of Voltage Application

The effect of the duration of voltage application on the breakdown voltage of treated oil was examined with direct, alternating and impulse voltages. For alternating and direct voltages three tests were made in which the rate of voltage application was varied so that breakdown occurred 10, 30 and 60 sec, respectively, after voltage application. In each case twenty breakdown voltages were determined after a number of conditioning discharges had been passed through the gap. Chromium-plated spheres of 13 mm diameter at a spacing of 1·25 mm were used and the results are given in Table 1. It may be concluded from these results that, for the times considered, the electric strength for both direct and alternating voltages is independent of the duration of the voltage application and, moreover, that the electric strengths are almost the same for the two types of voltage. This is evidence of the purity of the test sample and indicates the absence of certain impurities which often give rise to thermal breakdown in insulating oils.

Table 1

Effect of Duration of Voltage Application

Duration	Electric strength	
	A.C.	D.C.
sec	kV/cm	kV/cm
10	738	747
30	720	735
60	730	743

For the impulse tests on treated oil, impulses having shapes of 1/3, 1/5, 1/10 and 1/25 microsec were first used and tests with impulses of 1/50, 1/75 and 1/100 microsec were made subsequently. The electrodes were chromium-plated spheres of 13 mm diameter and the gap setting was 0·1 mm. The results of these tests are shown in Fig. 2, which, for the sake of comparison, also contains the results of untreated oil. The oil in this case was merely filtered at atmospheric pressure and room temperature through a sintered glass filter of grade No. 2. It may be assumed

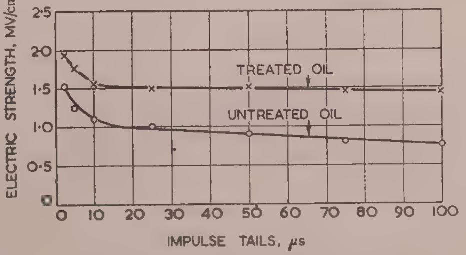


Fig. 2.—Effect of impulse duration.

that the filtration did not appreciably change the gas content of the oil. Each point on the curve in the figure represents the mean of from 5 to 10 individual measurements. It was observed during the tests that breakdown of the oil always occurred on the tail of the impulse wave.

It can be seen from Fig. 2 that for the treated oil the electric strength is almost independent of the duration of the impulse voltage, except for impulses having tails of less than 15 microsec, which produce a marked increase. For untreated oil, on the other hand, the electric strength is not only much smaller but it steadily increases as the duration of the impulse is reduced, this increase becoming more rapid for impulses having tails of less than 20 microsec. This would suggest that the breakdown of the untreated oil involves a mechanism which is time-dependent.

(3.2) Effect of Gap Setting

For these tests chromium-plated spheres of 13 mm diameter were used. The breakdown values were determined for spacings of 25, 50, 100, 200 and 500 × 10^{-3} mm. After conditioning, ten measurements with direct voltages of positive polarity were made for each of these spacings. The breakdown voltages for the first three spacings were measured with an electrostatic voltmeter connected directly across the test cell. The results are given in Table 2, which shows that the electric strength is constant in the range of 0·5–0·1 mm, but for shorter gaps there is a tendency for the electric strength to increase with decreasing spacing.

Table 2

Effect of Gap Setting

Gap setting	Electric strength	
	Direct voltage	Impulse voltage
mm × 10^{-3}	kV/cm	kV/cm
25	780	4 000
50	775	2 430
100	770	1 910
200	770	1 910
500	770	1 910

The results for impulse voltages of positive polarity having a waveshape of 1/3 microsec, and using the same electrodes, are also shown in Table 2, where each value represents the average of five measurements. As can be seen, the electric strength remains constant for spacings down to 0·1 mm and increases for smaller spacings.

For impulses of this short duration, and for small gaps, the electric strengths were very high. For example, for a gap setting of 15 × 10^{-3} mm the electric strength obtained was about 6 000 kV/cm. Moreover, since breakdown was observed to occur approximately 1 microsec after the impulse attained its peak value, it is to be expected that even higher values would be obtained for this gap setting with impulses of shorter duration.

(3.3) Effect of Electrode Shape

(3.3.1) Uniform Field Configuration.

The tests in uniform fields were made with spherical brass electrodes having diameters of 0·25 and 0·5 in, respectively, and with hemispherical electrodes of 1 in diameter. In addition, two sizes of plate electrodes, having a profile recommended by Bruce,[2] were used. They were made of brass and had diameters of 1·0 and 1·5 in respectively. The test voltage was a direct voltage of positive polarity; it was raised at such a rate that breakdown occurred from 10 to 15 sec after the instant of application and the time between successive breakdowns was approximately five minutes. After conditioning with alternating voltage, 50 breakdown measurements at a gap setting of 0·25 mm were made with each type of electrode. The results are given in Table 3.

Table 3

Effect of Electrode Shape

Electrodes	Electric strength			Standard deviation	Coefficient of variation
	Mean	Min.	Max.		
	kV/cm	kV/cm	kV/cm	kV/cm	%
Spheres:					
0·25 in diam. ..	716	667	820	32·4	4·5
0·50 in diam. ..	694	622	754	27·3	3·9
Hemispheres:					
1·0 in diam. ..	683	630	809	34·8	5·1
Plates:					
1·0 in diam. ..	668	594	725	29·2	4·3
1·5 in diam. ..	649	612	713	20·5	3·2

The Table shows that there is a slight tendency for the electric strength to increase with decreasing size of the electrodes but that this dependence is not very marked. This is in contrast with all previously published tests on insulating oils, in which a pronounced relation was usually observed, e.g. Bredner[5] and Clark.[6]

(3.3.2) Non-Uniform Field Configuration.

In the tests in non-uniform fields brass point-plane electrodes were used. The point electrode consisted of a needle of $\frac{1}{16}$ in diameter rod tapering uniformly to a point over a length of $\frac{1}{4}$ in. The gap setting was 1 mm and the measurements were made with alternating voltages, direct voltages of both polarities and impulse voltages. For the conditioning with electrodes of this type never more than 5 preliminary discharges were passed through the gap, but an increase of the breakdown voltage with successive discharges was apparent only when the point electrode was positive.

For direct and alternating voltages the electric strength was found to be completely independent both of the duration of the applied voltage and of the time interval between successive breakdowns. It was possible, for example, to obtain the same value for the breakdown merely by slightly reducing the test voltage and immediately raising it again. The measurements were very consistent, the maximum deviation being never more than 5%, and for a spacing of 2 mm the breakdown voltage could be reproduced to within 1 kV. A similar consistency of the measurements with non-uniform fields was reported by Dornte[7] and Lewis.[8]

With alternating and direct voltages of positive polarity breakdown occurred in the form of intermittent discharges which quickly quenched themselves and were accompanied by sharp clicking noises. It was difficult to assess the breakdown voltage in these cases and it was therefore decided to take the voltage readings during the occurrence of the intermittent discharges at regular intervals of about 1 sec.

During these tests, the discharge current through the test cell was not by-passed after breakdown of the oil. This, however, did not produce any noticeable decomposition of the oil, probably on account of the very short duration of the discharges. Also, changing the current-limiting resistance in the h.v. side of the test cell from 18 to 6 megohms had no effect on the measured breakdown values, nor did it affect the time sequence of the intermittent discharges. Table 4 gives the average value of

Table 4

Breakdown Voltage for Non-uniform Field

	Breakdown voltage
	kV
Alternating	25·4
Direct + ...	25·5
Direct −	31·0

10 measurements which were obtained with alternating and direct voltages. The results show, as was to be expected, since the breakdown voltage was independent of the time of voltage application, that the peak value of the alternating breakdown voltage was the same as the lower of the two direct voltages.

As can be seen, there is a pronounced polarity effect, the breakdown voltages for the negative point being higher than those for the positive point.

Similar results were obtained by Dornte[7] when testing dehydrated heptane with non-uniform electrode gaps. With air-saturated samples he found higher breakdown values when the polarity was positive, whilst carefully degassed samples gave higher values when the direct voltage was negative. Dornte also noticed intermittent discharges with both samples, and similar effects were observed by Sorge,[9] Schroter[10] and Race.[3]

For the impulse tests, impulses having a shape of 1/3 microsec were used. For each gap five measurements were made with each polarity; the average breakdown values are shown in Fig. 3.

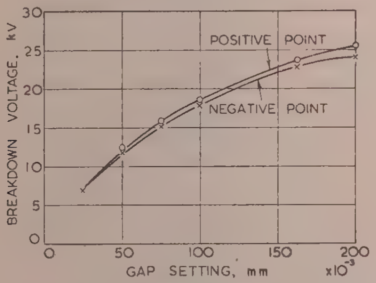


Fig. 3.—Polarity effect for impulse voltages and non-uniform fields.

From this Figure it can be seen that although the difference in the electric strength for the two polarities is small (less than 5%) it is nevertheless significant. Unlike the results for direct voltages, the breakdown values obtained with positive impulses are higher than those with negative impulses. This important difference in the polarity effect is discussed in Section 4, and in Section 3.8 it is shown how this effect is affected by the presence of foreign particles in the treated oil.

(3.4) Effect of Electrode Metal

The effect of the electrode metal on the breakdown voltage was examined for both uniform and non-uniform field configurations. For the tests with direct voltages and uniform fields, spherical electrodes of 13 mm diameter were used. The metals examined were chromium, silver, aluminium, stainless steel, steel, copper and brass. The electrodes were cleaned as described in Section 2.3 and the breakdown voltages were measured after completion of the conditioning process. It was found that conditioning varied according to the metal. For example, for chromium electrodes 13 preliminary discharges were needed, whilst for silver electrodes 30 discharges were necessary to obtain breakdown values which showed no tendency to increase further. The tests were made with direct voltage of positive polarity and the gap setting was 0·25 mm. For each metal 30 measurements were made; the results are shown in Table 5. This Table also contains the photo-electric work function of the corresponding non-degassed metals, as given by von Engel and Steenbeck. The work function for chromium is that given by Kösters.[11]

Table 5

Effect of Electrode Metal (Direct Voltage)

Electrode metal	Photo-electric work function	Electric strength			Standard deviation	Coefficient of variation
		Mean	Min.	Max.		
	eV	kV/cm	kV/cm	kV/cm	kV/cm	%
Chromium . .	4·7	742	662	816	29·3	3·9
Silver	3·09	748	631	857	44·8	5·9
Aluminium . .	1·77	844	790	927	43·2	5·1
Stainless steel	—	727	665	782	30·5	4·2
Steel	3·92	926	868	1 044	37·2	4·0
Steel, degassed	—	1 360	1 244	1 395	48·6	3·6
Copper . .	3·89	992	919	1 095	32·2	3·2
Brass	—	694	622	754	27·3	3·9

The Table shows that there is a marked dependence of the electric strength on the metal of which the electrodes are made. Copper, for example, gave an electric strength about 40% higher than that for brass. As is to be expected, there appears to be no relation between the electric strengths obtained with different metals and the corresponding work functions. These results will be discussed in Section 4.

To show the effect of the treatment of the electrode surface on the electric strength, Table 5 also contains the results which were obtained with steel electrodes after they had been degassed. The degassing was carried out for three hours at a pressure of 10^{-3} mm Hg, during which time the electrodes were heated by means of a high-frequency coil and after which they were left to cool for three hours under the same vacuum. For this test, in order to avoid the introduction of air, the test cell was not rinsed with the test sample. After conditioning with 15 alternating voltage discharges, 50 breakdown values were measured and from these the values given in the Table were determined.

For the impulse tests with uniform field configurations, spherical electrodes of 13 mm diameter were used; the combinations examined were steel–steel, aluminium–aluminium, and chromium–steel.

Electrodes of the same metal were tested with 1/3 microsec impulses of positive polarity, whilst impulses of the same shape and both polarities were used for the electrodes of different metals. Each combination was tested at gap settings of 50, 100, 200 and 300 × 10^{-3} mm, respectively, and for each gap and polarity of the impulse voltage five breakdown values were determined. The results are shown in Table 6, which, for comparison, also contains the results for the chromium-plated spheres given in Section 3.2.

As can be seen, the difference in the electric strength values obtained with electrodes of different metals is small and of the order of magnitude of the experimental error. This is in striking contrast to the results with direct voltages, where a pronounced dependence was found. It would appear, therefore, that the dependence on the electrode metal is a time-dependent effect. This seems to be confirmed by the results given by Watson and Higham[16] in their work on the effect of degassing of the test electrodes. Using impulses of 4 millisec duration they obtained

Table 6

EFFECT OF ELECTRODE METAL (IMPULSE VOLTAGE)

Electrode metal	Electric strength			
	50 × 10^{-3} mm gap	100 × 10^{-3} mm gap	200 × 10^{-3} mm gap	300 × 10^{-3} mm gap
	MV/cm	MV/cm	MV/cm	MV/cm
Steel–steel	2·60	1·96	1·96	1·96
Aluminium–aluminium ..	2·43	1·99	1·99	1·99
Chromium (cathode) ..	2·48	1·88	1·90	1·92
Steel (cathode)	2·48	1·88	1·90	1·92
Chromium–chromium ..	2·43	1·91	1·91	1·91

electric strength values for electrodes made of phosphor-bronze and steel respectively, which showed a percentage difference of 32%, whereas the percentage difference in the electric strength for these electrodes was only 2·8% when impulses of 0·2 millisec duration were used.

The experiments with non-uniform field configurations were made with point-plane electrodes. The metals examined were chromium, aluminium and copper, as well as a combination consisting of a steel point and a chromium plate. The gap setting throughout these experiments was 1 mm and direct voltages were used. For every combination ten measurements were made with positive polarity followed by ten measurements with negative polarity. The by-passing switch across the test cell was disconnected during these tests.

The breakdown voltages obtained were almost identical with those given in Section 3.3.2 for brass point electrodes. With the point positive, the maximum difference recorded was less than 5%, and identical results were obtained when the point was negative.

(3.5) Effect of Irradiation

In this Section the results are given of the effect of irradiation on the breakdown voltage of treated oil. Cobalt-60 γ-ray sources of 0·6, 0·8 and 15 millicurie were used. The 15 mc source was in a thick-walled container which had a window 2 in deep. The two other sources were in thin-walled aluminium tins. For irradiation, the 15 mc source was placed with its window facing the electrode gap and at a distance of about 1 in from the wall of the test cell and the other sources were placed in contact with it. All sources were used at the same level as the electrode gap.

In all experiments with direct voltages in which irradiation was used, the flow of an appreciable conduction current through the oil was observed. Since the test cell was not provided with guard rings it is not possible to say whether the conduction current caused by the irradiation was due to charges in the bulk of the liquid or to charges on the glass walls of the test cell. It should be mentioned, however, that in the course of a previous investigation[19] irradiation tests on pure simple organic liquids were made in the laboratory under similar conditions. These tests did not show an increase of conduction current, and it would therefore appear that for treated oil irradiation does give rise to charge carriers in the liquid.

Chromium-plated spherical electrodes of 13 mm diameter at a gap setting of 0·5 mm were used. After the conditioning process, the gap was irradiated with the 0·8 mc source and 50 breakdown voltages were determined with direct voltage of positive polarity. The rate of voltage rise was such that breakdown occurred 15–20 sec after application of voltage.

The results of these tests are given in Table 7. A comparison of the electric strength values shows that although the difference is small it is of a magnitude which cannot be accounted for by the scatter of individual measurements.

Table 7

EFFECT OF IRRADIATION

	Electric strength			Standard deviation	Coefficient of variation
	Mean	Min.	Max.		
	kV/cm	kV/cm	kV/cm	kV/cm	%
With irradiation ..	709	674	778	23·8	3·28
Without irradiation ..	742	662	816	29·3	3·9

For the impulse voltages the same electrodes were used. After conditioning with alternating voltages at a spacing of 0·25 mm, the breakdown values were determined with 1/3 microsec impulses of positive polarity for spacings of 100, 200 and 300 × 10^{-3} mm. For each spacing ten measurements were made without irradiation followed by ten measurements with the gap irradiated by all three cobalt sources simultaneously. No difference in the electric strengths could be detected in the two cases and it was therefore concluded that irradiation of the gap with γ-rays had no effect on the impulse breakdown voltage of the treated oil.

(3.6) Effect of Temperature

The test cell was immersed in an oil bath heated by a small immersion heater with means for regulating the temperature to within ±1·5° C. The temperature was measured by a mercury thermometer placed in the oil bath.

After the conditioning process, ten measurements were made at each of the following temperatures: 18 (room temperature), 50, 70, 80, 90 and 100° C. At each temperature a period of twenty minutes was allowed before the measurements were begun in order to ensure equalization of the temperature of the oil-bath and test liquid and the stabilization of the sample at that temperature. The tests were made with direct voltages of positive polarity and with a gap setting of 0·25 mm. Correction was made for the variation of the gap setting with temperature, which amounted to 0·347 micron/°C. This correction was previously determined experimentally by finding the temperature required to reduce to zero a gap setting which at room temperature was 25 microns.

The results are given in Table 8, and it can be seen that there is no tendency for the electric strength to increase with increasing temperature in the range from 18 to 50° C. This is in contrast to the findings of most previous investigators, e.g. Clark,[12] Hoover and Hixon[13] and Lazarev.[14]

Table 8

EFFECT OF TEMPERATURE

Temperature	Electric strength
°C	kV/cm
18	740
50	730
70	735
80	725
90	710
100	675

Breakdown measurements on treated oil were also made at temperatures below 18° C. Although these measurements were not wholly satisfactory, mainly because of the difficulty of maintaining the temperature long enough for equilibrium between sample and cooling bath to be established, they nevertheless showed that there was no tendency for the electric strength to vary at temperatures down to −35° C. Below this temperature a steep rise in the electric strength could be observed.

(3.7) Effect of Pressure

The effect of external pressure on the electric strength of treated oil was examined for uniform and non-uniform field configurations. In these tests the test cell was completely filled and sealed off by means of mercury in the side tube (see Section 2.2.3) before placing it in the pressure vessel. Both direct and alternating voltages were used and the test voltage was raised at such a rate that breakdown occurred between 15 and 20 sec after application of the voltage; intervals of 10 min were allowed between successive breakdowns.

For uniform fields, chromium-plated spherical electrodes of 13 mm diameter at a spacing of 0·25 mm were used. The conditioning process was carried out at atmospheric pressure and the subsequent measurements were made at pressures of 25, 50, 75, 100 and 200 lb/in^2. The results are given in Table 9, in which each electric strength represents the mean of ten measurements.

Table 9

Effect of Pressure

Gauge pressure	Alternating voltage	Direct voltage
lb/in^2	kV/cm	kV/cm
0 (atm)	725	732
25	770	776
50	750	780
75	750	766
100	765	770
200	760	778

The table shows that, although the smallest electric strength was obtained at atmospheric pressure, the difference in the values for different pressures is not very pronounced. For insulating oil this result is new and is, no doubt, due to the careful treatment which the oil received before the test.

For the non-uniform field, chromium-plated point-plane electrodes were used. The test procedure was exactly the same as for the uniform field. After conditioning, measurements were made with direct voltages of both polarities at pressures of 0 (atmospheric), 10, 25, 50, 100, 150 and 200 lb/in^2. The gap setting was 0·6 mm; the results are shown in Fig. 4, where each point represents the mean of five measurements. It can be seen from this Figure that a pressure-dependence of the breakdown voltage, which was not very marked, was observed only when the point was negative. The positive breakdown values, which were lower, did not change with external pressure. It was again found, with positive polarity, that the breakdown was of the intermittent type, as described in Section 3.3.2. For the negative polarity this type of discharge was sometimes observed when the external pressure exceeded 50 lb/in^2.

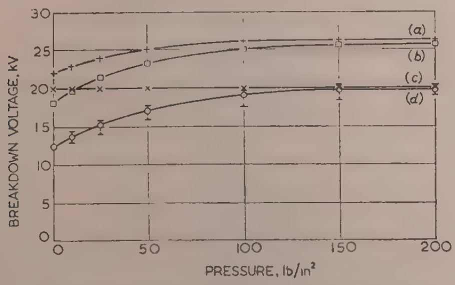


Fig. 4.—Effect of pressure, non-uniform field.

(*a*) Negative polarity, treated oil.
(*b*) Negative polarity, oil containing some gas.
(*c*) Positive polarity, treated oil.
(*d*) Positive polarity, oil containing some gas.

Since the treated oil used for these tests was carefully degassed, it was considered instructive to repeat them with oil which contained some gas. Accordingly, a sample of treated oil was exposed for 15 min to dry and filtered air before being tested. At each pressure and after every breakdown an interval of twenty minutes was allowed to elapse before the next voltage application. Hoover and Hixon[13] also found that intervals of similar duration were necessary for the stabilization of the measurements with samples containing dissolved gases.

The results of these measurements are also shown in Fig. 4, where each point represents the mean of five measurements. During these tests the breakdown voltages for the positive polarity were difficult to determine, particularly at pressures near 100 lb/in^2. The range of voltage over which intermittent discharges occurred was much wider than for treated oil, and in Fig. 4 the upper and lower limits for the onset of these discharges are indicated by dotted lines. It will be noted that the presence of the gas in the oil increases the pressure-dependence of the breakdown voltage now found for both polarities of the test voltage. As the pressure increases, the upper limits of the breakdown voltage tend to assume values almost identical with those obtained for treated oil.

(3.8) Treated Oil with Added Impurities

Some experiments were made on treated oil to which different impurities were added. First, the effect of adding water will be described. The tests were made with alternating and direct voltages of positive polarity on samples having water contents of 30, 50, 100, 150, 200 and 500 parts in 10^6.

To prepare the test samples an amount, corresponding to the above concentrations, of pure water was placed in a glass flask of 500 cm^3 capacity which had previously been carefully cleaned and dried. This flask was filled with treated oil from the collecting flask of the test apparatus through a glass tube reaching almost to the bottom of the flask. During the filling of the flask the oil was in contact with the atmosphere for about 2 or 3 min, so that it could be assumed that the amount of air absorbed by the oil was small and its effect on the electric strength could be neglected in comparison with the effect of the added water.

The flask, with the oil, was then put in an oil heating bath and, except for occasional shakings, was left there for two hours at a temperature of 73° C. According to Clark[15] the solubility of water increases considerably at this temperature and the small amounts of water that were added to the treated sample were easily dissolved, as could be seen from the clear appearance of the oil when examined with suitable illumination. At this stage the flask was allowed to cool slowly. Slow and gradual cooling is essential in order to obtain a water-in-oil emulsion with the smallest possible size of water particles. The sample was left for about two hours to ensure stabilization of the equilibrium between the emulsion and molecular phases of the water at the prevailing conditions. Before filling the test cell it was rinsed with part of the sample.

The tests were made with a gap setting of 1 mm between chromium-plated spheres of 13 mm diameter, and for each water-content thirty individual measurements were made, with intervals of 10 min between successive breakdowns.

In these tests it was observed that there was a strong dependence of the breakdown voltage on the time of voltage application, especially with samples containing more than 30 parts in 10^6 of water, and the normal procedure of the one-minute test was adopted. The results for both alternating and direct voltage measurements are shown in Fig. 5, from which it can be seen that the addition of even small amounts of water to treated oil caused a considerable reduction in its electric strength. The shape of the curve of variation of electric strength with water content is similar to those given by previous investigators. There is, however, no quantitative agreement, as is to be expected in view of the differences in the oil used and in the treatment given to it before it was tested.

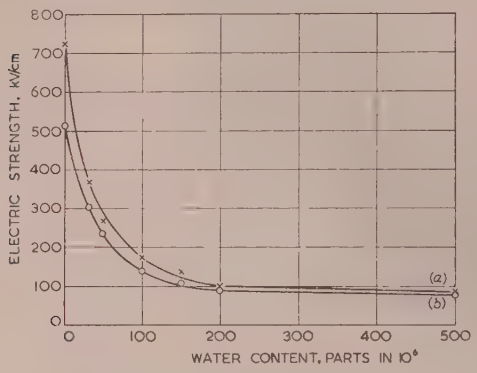


Fig. 5.—Effect of water content.

(a) Direct voltage.
(b) Alternating voltage.

During the experiments with direct voltages, measurements were also made with different gap settings between 0·5 and 2·5 mm, and it was found that for this range of spacings the breakdown voltage varied linearly with the gap.

With a gap setting of about 2 mm, it was possible after breakdown to observe two kinds of bubbles in the test cell. A small number of one kind, which were distributed throughout the liquid, slowly rose to the top of the cell. Those of the other kind, which seemed more numerous (their number seemed to vary with the water-content of the sample), were observed to oscillate between the electrodes along paths which followed the lines of force of the electric field; their velocity decreased as the voltage across the electrodes was reduced. Occasionally two of these bubbles, coming from different electrodes, were seen to collide, and this was accompanied by a minute visible spark. This second type of bubble probably consisted of water vapour formed during breakdown. This observation would suggest, therefore, that with undried oil conditioning might be partly due to gradual evaporation of its water content.

(3.9) Effect of Water on Suspended Matter

It has been known for some time that the effect of water on the electric strength is particularly harmful if the liquid dielectric contains suspended matter. This is explained by assuming that the water is adsorbed by the foreign matter (fibres or other particles) so that conducting filaments are formed which may bridge the gap. No explanation, however, has been found for the effect of water when the liquid contains suspended matter of extremely small (e.g. colloidal) size, when the formation of bridges seems unlikely.

Accordingly, a number of simple observation tests were made, using 100% spherical glass particles of 100-microns diameter. Although these particles were not very small, bridge formatio was considered unlikely on account of their shape.

A small amount of these glass particles, previously washed with acetone and distilled water and dried in an oven for two hours at 110° C, was introduced between 1 in-diameter plate electrodes in a sample of treated oil. Most of the particles rested on the surface of the lower electrode. No effect on the particles was observed when a direct voltage of 20 kV and positive polarity was applied across the gap, which was about 4 mm. However, on reversing the polarity the particles were attracted by the negative electrode, as would be expected from Cohen's law, but no further motion could be seen.

When the experiment was repeated with an oil sample containing water to about 80 parts in 10^6, most of the particles were observed to oscillate between the electrodes. This occurred for both polarities of the applied voltage. With oil samples of commercial purity, some only of the spheres oscillated whilst the rest remained attached to the negative electrode.

It remains now to describe experiments showing how the polarity effect for point-plane electrodes was affected by the presence of impurities in the oil. For this work the following samples were prepared:

(a) Treated oil to which was added a certain number of 100% spherical glass particles of 100 microns diameter.

(b) Untreated oil containing damp spherical polystyrene particles of 10 microns diameter.

(c) Dried oil saturated with gas.

In the experiments with the glass particles five breakdown values of the treated oil were first determined at a gap setting of 1 mm and with direct voltage of both polarities. Next, a small quantity of cleaned and carefully dried glass particles was introduced into the test cell and a further five measurements were made with each polarity. From the results, which are given in Table 10, it can be seen that the effect of adding the particles was appreciable for the negative point but that the breakdown voltage was not affected when the point was positive.

Table 10

Polarity Effect for Impure Oil

Sample	Breakdown voltage	
	Positive point	Negative point
	kV	kV
Treated oil	25·0	31·0
Treated oil with glass particles ..	25·0	28·5
Untreated oil	21·0	24·0
Untreated oil containing 3 cm³ of colloidal solution	18·0	19·5
Untreated oil containing 8 cm³ of colloidal solution	12·0	11·0

For the tests with samples containing damp polystyrene particles, a temporarily stable solution of polystyrene in oil wa prepared by adding about 0·5 g of polystyrene particles t 30 cm³ of untreated oil. A measure of 3 cm³ of this solution wa then added to about 200 cm³ of untreated oil and shaken Measurements were made immediately, using a gap setting o 1 mm, and for each polarity three breakdown values were determined. After the ninth measurement a rise in the breakdow voltage was observed, probably due to the removal of polystyren particles from the electrode gap, the gradual evaporation of th water-content and the precipitation of the unstable colloida particles.

The same test was repeated with an untreated sample to which 8 cm^3 of the colloidal solution was added. Results of these tests, together with the results for the untreated oil, are given in Table 10. It can be seen that the addition of some damp solid particles removes the difference in the breakdown voltages for the two polarities. No significance can be attached to the higher breakdown value for the positive polarity when 8 cm^3 of the colloidal solution was added to the sample, since the values in the Table are the averages of a few measurements only.

For the test on dried oil saturated with gas, the oil sample was dried in the presence of sodium wire and self-indicating silica gel of mesh size 20, for more than four weeks. It was then filtered gently through a sintered glass filter of grade No. 2 at room temperature and atmospheric pressure. In this way the sample retained its gas content and a gentle shaking was sufficient to produce a considerable cloud of gas bubbles, which probably consisted of hydrogen. It must be assumed that the dry sample, in addition to the gas, contained also a certain amount of suspended matter, but it was considered that the high gas content would be the dominant factor in the breakdown process.

Five measurements were made at each polarity of the direct voltage and at each of three gap settings—0·5, 1 and 2 mm. The scatter of the individual breakdown voltages was much greater than for treated oil, particularly when the point was negative. A possible explanation for this might be the presence of suspended particles. According to Cohen's law, such particles will be positively charged and hence will disperse through the liquid in the vicinity of the negative point electrode and thus cause scatter of the measurements.

In Fig. 6 the breakdown voltages are plotted against gap setting. For comparison, the Figure also contains the results for treated oil. The effect of the gas in the oil is considerable; it brings about an increase in the breakdown voltage when the point is negative and a decrease when the point is positive.

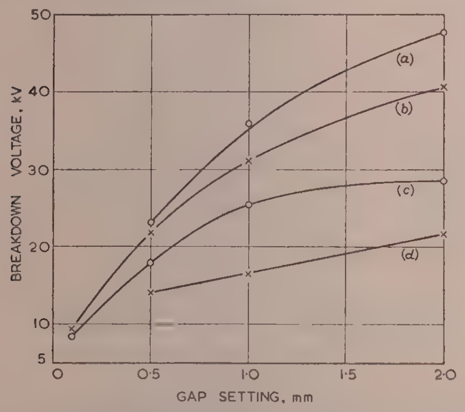


Fig. 6.—Change of polarity effect with gas content in non-uniform field.

(*a*) Negative polarity, oil saturated with gas.
(*b*) Negative polarity, treated oil.
(*c*) Positive polarity, treated oil.
(*d*) Positive polarity, oil saturated with gas.

(4) DISCUSSION OF RESULTS

The high alternating and direct breakdown values which were obtained for the treated oil and their independence of the duration of the applied voltage are an indication that this oil was fairly free from moisture and foreign particles, for the presence of such impurities would have caused breakdown at much lower voltages. With oil of this purity the breakdown is mainly affected by its gas content.

The dependence of the electric strength of mineral oil on its gas content was examined by Clark, who showed that, as the gas content of highly degassed oil is increased, the electric strength falls rapidly at first and, after reaching a minimum, rises again as the gas content is further increased. This remarkable behaviour of oil was confirmed during the work described here. Another result which is important has been reported by Walther and Tscheljuskina.[17] These workers found that the breakdown strength of toluene was independent of pressure when the toluene contained dissolved gas only, while the presence of the gas in suspension resulted in a pressure-dependence.

In the experiments with treated oil a marked reduction of the electric strength was noticed when its gas content was increased, for example, by exposing the oil for a short time to the atmosphere. In the light of Clark's result it would follow, therefore, that the treated oil contained very little gas and, in this respect, it corresponded to the highly degassed oil of Clark. Moreover, as shown in Section 3.7, the pressure-dependence of the electric strength of treated oil was small, which, from the results of Walther and Tscheljuskina, would indicate that the amount of gas in suspension was small.

The pressure-dependence of the breakdown voltage is usually taken to indicate that the breakdown in the liquid is of the gaseous type. The main features of the experimental results can be described qualitatively by the usual assumption that the breakdown takes place in gas bubbles in the oil or in its vapour phase. It is assumed, however, that two processes are involved in the initiation of the breakdown. The first consists in the formation of gas bubbles in the liquid at the cathode by the strong breakdown field and the introduction of electrons into the liquid by the ionization of these bubbles. The second process is an adequate electron multiplication by a gaseous ionization in the bulk of the liquid.

Under suitable experimental conditions the formation of bubbles on the electrodes can be observed. It will depend on the magnitude of the electric field and on the external pressure and will be greatly influenced by the amount of gas available at or near the electrodes, their surface condition and the metal of which they are made. It will be of a statistical nature and the probability of its occurrence will increase with the size of the electrodes.

The nature of the second process need not be specified for the present purpose, but, whatever it is, this process will be affected by the positive space charge which is formed by the ionization. It is reasonable to assume that the space-charge formation will depend on the gas content of the liquid, for, even if ionization occurs in the vapour phase, the vapour formation will take place at the boundary between the liquid and the gas bubbles, which act as nuclei. Hence, as the gas content increases more space charge will form near the cathode and this will result in a change in the field distribution in the electrode gap. The field will increase near the cathode and it will be weakened in the remainder of the gap. The space charge will therefore tend to confine the ionization to the cathode region and a higher voltage will be required to extend it towards the anode. Thus, whereas the conditions for the two processes become more favourable as the gas content of the liquid is increased, beyond a certain gas content higher voltages will be required for the second process because of the retarding effect of the positive space charge.

This is important, for it explains Clark's result mentioned earlier. With fields that do not depart appreciably from the uniform field, experimental evidence seems to show that for treated oil the voltages required for the two processes are about the same. The first process will trigger the discharge, and a marked dependence of the breakdown voltage on the metal and the surface condition of the electrode is to be expected; this was found in the experiments described in Section 3.4. Moreover, as already mentioned, this process will occur more readily as the effective area of the electrode is increased, which may explain the slight decrease of electric strength with increasing size of electrodes, which is shown in the results of Section 3.3.

With gas contents beyond a certain value the breakdown will be determined by the second process, which will require the higher voltage of the two. In this case no electrode effect of the breakdown voltage should be expected. This may explain the conclusions of earlier investigators, who do not report any dependence of the electric strength on the electrode metal. Most of this work was done with open test cells, and, provided that the oil was sufficiently free from moisture and other foreign particles which would cause breakdown by a different mechanism at low voltages, it is likely that it contained an appreciable amount of gas.

The dependence of the electric strength on the gas content also explains the results on the conditioning of treated oil. As stated in Section 2.4 the electric strength of dried and filtered oil, subjected to a few degassing cycles only, indicated a pronounced conditioning, and its breakdown values increased with each successive discharge during about 50 discharges, which was the number usually observed. With treated oil (i.e. 10 degassing cycles) this increase was found only during a relatively small number of initial discharges. The reason for this is that every discharge in the liquid is accompanied by a small evolution of gas, and the resulting reduction of the gas content causes an increase in the breakdown strength. With treated oil a gas content is ultimately reached at which, to satisfy the conditions of equilibrium, the gas liberated at each discharge is reabsorbed by the oil and conditioning ceases. This reabsorption of the gas by the liquid would explain why, at this stage, gas bubbles produced by the discharges do not appear to affect the breakdown, as has been observed also by other investigators.

An experimental observation which may have a bearing on this conclusion should be mentioned here. As a rule, with oil which has not been properly degassed, gas bubbles can be seen to rise to the top of the test cell after each discharge, but with degassed oil it is sometimes possible, after conditioning, to observe how a gas bubble freed by the discharge disappears in the bulk of the liquid.

From Clark's result it will be appreciated that for liquids of sufficiently high gas content the conditioning effect may be reversed, i.e. the electric strength may decrease with successive discharges, as has been found by a number of other investigators. For oil which has not been dried and filtered, conditioning which is sometimes observed is very probably due to the gradual removal of dust particles and to drying by the discharges.

When the field distribution departs widely from the uniform distribution, as, for example, for point-plane electrodes, the two processes will require different voltages according to the polarity of the test voltage, and the breakdown voltage will be determined by one or the other of them. For the negative point, bubble formation will occur at this electrode first as the applied voltage is raised, and a higher voltage will be necessary for the second process, i.e. the ionization in the liquid. Moreover, the effect of the positive space charge in confining the ionization to the vicinity of the negative point will be greater for this field distribution than for the uniform field and will increase with the gas content of the liquid. The second process, modified by the presence of the space charge, will therefore determine the breakdown voltage for this polarity. This voltage should increase with gas content, as was found in the experiments (see Section 3.9).

For the positive polarity the electric field assumes its smallest value at the plate cathode. When the field at this electrode reaches a value sufficient for the first process, the second process will also take place, since the field in the liquid will be higher. For this polarity, therefore, the breakdown voltage will be dete mined by the first process and, since the chance of bubb formation increases with the gas content of the liquid, a reductic in the breakdown voltage with gas content is to be expected; th is confirmed by the experiments of Section 3.9.

The conditions for ionization will also be different wi positive polarity. The electrons liberated by the first proce will move in an electric field which rapidly increases as th positive point is approached. The ionization will set up positiv space charge near this electrode, with the result that the field a the plate will be enhanced and the supply of electrons from th first process will increase rapidly. Thus, whereas for the negativ point the space charge exerts a stabilizing effect on the discharg with positive polarity it leads to unstable conditions. The fiel at the plate will quickly assume very high values so that a copiou stream of electrons will leave the plate, and it is suggested tha in moving towards the other electrode they will neutralize th positive space charge. This would lead to the original fiel distribution in the gap and a quenching of the discharge. It believed that this sudden instability constitutes the intermitter type of discharge observed during the experiments. Since it triggered by the first process, the random nature of this proce would explain its irregular occurrence, which was very noticeab and which increased with the gas content.

Since for a given applied voltage the electric field at the poi electrode is greater than at the plate, a higher voltage will b required for bubble formation at this electrode than at the poin However, the difference in voltage need not be as great as migl be expected from the field distribution, for, as mentioned alread bubble formation also depends on the effective area of th electrode and the volume of the adjacent liquid. It is very likel therefore, that a bubble might form at a propitious point on th plate at a much lower electric field than would be required for th point electrode. But, even if bubble formation at the negativ point occurs at a lower voltage, the breakdown voltage for th polarity may exceed that for the positive polarity, as found i the experiments, since it is governed by the second process an by the retarding effect of the space charge on this process.

Both processes suggested for the breakdown involve gaseou ionization and they should therefore be affected by the extern pressure, so that the breakdown voltage should be pressur dependent. This was found in the experiments with the unifor and non-uniform electrode configurations. The pressur dependence for treated oil was small in both cases, and tests wi non-uniform fields showed that it increased with the gas conte of the oil. For both field distributions the pressure-dependen decreased with increasing pressure, and for sufficiently hi pressures the breakdown voltage approached a constant valu In the light of the result of Walther and Tscheljuskina it wou seem reasonable to suppose that as the external pressure is i creased the gas suspended in the oil is gradually dissolve in accordance with Henry's law, and the breakdown streng will become pressure-independent when all the gas is in solutic

To explain the pressure-independence of the electric streng at high pressures it has been suggested that ionization takes pl in the liquid phase. This does not seem probable from measured breakdown fields. Moreover, since much hig strengths are obtained for impulse voltages, it follows that process operative for alternating and direct voltage breakdo must be time-dependent. Ionization of the dissolved gas m cules is even more unlikely, since the ionization potential of th molecules, although lowered by the presence of the oil, still be greater than that of the oil molecules. Howe another process, which would be time-dependent and wh could operate at lower voltages, might occur. The elec fields which have been observed may be sufficiently high for electrons, in moving through the liquid, to overcome the binding forces of the dissolved gas molecules (which are probably of the Van der Waals type) and, by a process of nucleation, gas bubbles may form from the detached gas. The field required for this detaching process may be high enough for the ionization of the gas suspension, whatever the external pressure. Such a mechanism would render the two suggested breakdown processes independent of pressure and it would explain the absence of pressure-dependence found in the experiments on treated oil when tested with point-plane electrodes and direct voltages of positive polarity.

The two processes assumed for the direct-voltage breakdown of treated oil are time-dependent, and experiments using impulse voltages of microsecond duration have shown that they are not operative. With these short impulse voltages, treated oil gave electric strengths which are comparable with the highest reported for simple organic liquids. The results are also similar in several other respects; in both cases the metal of the electrode has little effect on the breakdown for uniform fields, and the breakdown voltage does not appear to be affected when the electrode gap is irradiated with γ-rays. More important is the agreement in the results on the polarity effect for point-plane electrodes. As for pure organic liquids, higher breakdown values were obtained for treated oil when the polarity was negative.

It would therefore appear that for impulse voltages the breakdown process of treated oil is akin to that occurring in pure liquids. The breakdown fields are so high that electron emission from the cathode and ionization in the liquid phase are possible. These two processes would then replace the two processes suggested for the direct-voltage breakdown. They were put forward by Lewis[18] for the breakdown of simple organic liquids and are capable of explaining reasonably a number of the results observed with these liquids.[19]

It may be concluded from the experimental work that breakdown of the transformer oil may be due to a number of different processes, and breakdown voltages may vary from the very low values given in Sections 3.8 and 3.9 for oil containing impurities to the very high values which were obtained with very short impulse voltages. For oil which has been carefully dried and filtered, the experimental evidence seems to indicate that the direct-voltage breakdown is governed by the gas content of the oil, and high electric strength values can be obtained for highly degassed oil.

The mode of breakdown which has been suggested for this type of liquid, although capable of explaining a number of the experimental observations, is over-simplified and incomplete in several important details. The experimental results suggest further tests, particularly with non-uniform field configurations, which may confirm and clarify the assumptions made. Important also in this connection is the study of the equilibrium between gas in suspension and in solution in the oil, the formation, at the electrodes and in the liquid, of gas bubbles from the dissolved gas molecules, and other related problems. In this work, which is being pursued, the recent views on nucleation may prove helpful.

(5) REFERENCES

(1) Polley, M. H.: "Hydrocarbon Absorption by Stopcock Lubricants," *Analytical Chemistry*, 1951, **23**, p. 545.

(2) Bruce, F. M.: "Calibration of Uniform Field Spark Gaps for High-Voltage Measurement at Power Frequencies," *Journal I.E.E.*, 1947, **94**, Part II, p. 138.

(3) Rage, H. H.: "Dielectric Strength of Insulating Liquids in a Continuously Circulating System," *Transactions of the American I.E.E.*, 1940, **59**, p. 730.

(4) Russel, A.: "The Dielectric Strength of Air," *Philosophical Magazine*, 1906, **11**, p. 237.

(5) Bredner, R.: "Dielektrische Festigkeiten und Verluste Flüssiger Kohlennasserstoffe mit und ohne Dipolecharakter," *Archiv für Elektrotechnik*, 1937, **31**, p. 351.

(6) Clark, F. M.: "The Role of Dissolved Gases in Determining the Behaviour of Mineral Insulating Oils," *Journal of the Franklin Institute*, 1933, **215**, p. 39.

(7) Dornte, R. W.: "The Dielectric Strength of Benzene and Heptane," *Journal of Applied Physics*, 1939, **10**, p. 514.

(8) Lewis, T. J.: "The Dependence of the Dielectric Strength of Pure Liquids on the Cathode Material," *Proceedings of the Physical Society*, B, 1953, **66**, p. 425.

(9) Sorge, J.: "Über die elektrische Festigkeit einiger flüssiger Dielektrika," *Archiv für Elektrotechnik*, 1924, **13**, p. 189.

(10) Schröter, S.: "Reinigung und Durchschlagfestigkeit von Transformato renoil," *ibid.*, 1923, **12**, p. 67.

(11) Kösters, H.: "Messungen von Voltaspannungen zwischen reinen Metallen," *Zeitschrift für Physik*, 1930, **66**, p. 807.

(12) Clark, F. M.: "Dielectric Strength of Mineral Oils," *Electrical Engineering*, 1935, **54**, p. 50.

(13) Hoover, W. G., and Hixon, W. A.: "Dielectric Strength of Oil with Variation of Pressure and Temperature," *Transactions of the American I.E.E.*, 1949, **68**, p. 1047.

(14) Lazarev, A., and Nigmatulina, L.: "The Equilibrium and Electric Strength of Two-Phase Water Liquid Dielectrics," *Journal of Technical Physics of the U.S.S.R.*, 1938, **5**, p. 195.

(15) Clark, F. M.: "Water Solution in High-Voltage Dielectric Liquids," *Transactions of the American I.E.E.*, 1940, **59**, p. 433.

(16) Watson, P. K., and Higham, J. B.: "Electric Breakdown of Transformer Oil," *Proceedings I.E.E.*, Paper No. 1501 M, March, 1953 (**100**, Part IIA, p. 168).

(17) Walther, A., and Tscheljustkina, O.: "Role of Gases in the Breakdown of Liquid Dielectrics," *Journal of Technical Physics of the U.S.S.R.*, 1936, **3**, p. 940.

(18) Lewis, T. J.: "Electrical Breakdown in Organic Liquids," *Proceedings I.E.E.*, Paper No. 1488 M, March, 1953 (**100**, Part IIA, p. 141).

(19) Maksiejewski, J. L., and Tropper, H.: "Some Factors affecting the Measurement of the Electric Strength of Organic Liquids," *ibid.*, Paper No. 1642 M, April, 1954 (**101**, Part II, p. 183).

521.3

Monograph No. 13
June 195

A SHORT TABLE OF THE LAGUERRE POLYNOMIALS

By LUCY J. SLATER, M.A., Ph.D.

(*The paper was first received* 31*st December*, 1954, *and in revised form* 1*st March*, 1955.
It was published as an Institution Monograph *in June*, 1955.)

In view of the interest aroused in Laguerre polynomials by a recent paper,[3] there is need for the publication of the short Table of the function $L_n(x)$ referred to therein. This Table is over the range

$$n = 0(1\cdot0)10\cdot0, \quad x = 0(0\cdot1)5\cdot0$$

and is correct to six places of decimals. Modified second differences are provided for interpolation[1] with

$$\delta_m^2 = \delta^2 - 0\cdot184\delta^4 + 0\cdot038082\delta^6$$

The Laguerre polynomial is a special case of the confluent hypergeometric function, and the Table was calculated by direct summation of the series

$$\mathrm{F}(-n;\,1;\,x) \equiv L_n(x) = \sum_{r=0}^{n}(-n)(-n+1)(-n+2)\ldots(-n+r-1)x^r/(r!)^2 \quad . \quad (1)$$

on the electronic calculator Edsac 1 in the Mathematical Laboratory of Cambridge University, as part of the preparatory work for a larger table of $\mathrm{F}(a;\,b;\,x)$ over the same range in x. The calculations were checked by differencing by hand in the x-direction and by the recurrence relation

$$(n+1)L_{n+1}(x) = (2n+1-x)L_n(x) - nL_{n-1}(x) \quad . \quad . \quad . \quad (2)$$

in the n-direction.

The result of eqn. (2) can also be used to extend the Table to other ranges of n, and the relation

$$L_{-n}(-x) = \varepsilon^{-x}L_{+n-1}(x) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

can be used to provide for an extension of the Table to negative values of x. Full details of the properties of the Laguerre polynomials will be found in Reference 2, Section 10.12.

REFERENCES

(1) British Association: "Auxiliary Tables No. I, Coefficients of Modified Everett Interpolation Formulae" (Cambridge University Press, 1946).

(2) Erdelyi, A.: "Higher Transcendental Functions, Vol. II" (Bateman Project, New York, 1953).

(3) Lampard, D. G.: "A New Method of Determining Correlation Functions of Stationary Time Series," *Proceedings I.E.E.*, Monograph No. 104 R, August, 1954 (**102** C, p. 35).

Correspondence on Monographs is invited for consideration with a view to publication.
Dr. Slater is at Newnham College, Cambridge.

x	$n = 0$	δ^2	$n = 1$	δ^2	$n = 2$	δ^2
0·0	1·0	0·0	1·0	0·0	1·000	0·01
0·1	1·0		0·9		0·805	1
0·2	1·0		0·8		0·620	1
0·3	1·0		0·7		0·445	1
0·4	1·0		0·6		0·280	1
0·5	1·0		0·5		0·125	0·01
0·6	1·0		0·4		−0·020	1
0·7	1·0		0·3		−0·155	1
0·8	1·0		0·2		−0·280	1
0·9	1·0		0·1		−0·395	1
1·0	1·0		0·0		−0·500	0·01
1·1	1·0		−0·1		−0·595	1
1·2	1·0		−0·2		−0·680	1
1·3	1·0		−0·3		−0·755	1
1·4	1·0		−0·4		−0·820	1
1·5	1·0		−0·5		−0·875	0·01
1·6	1·0		−0·6		−0·920	1
1·7	1·0		−0·7		−0·955	1
1·8	1·0		−0·8		−0·980	1
1·9	1·0		−0·9		−0·995	1
2·0	1·0		−1·0		−1·000	0·01
2·1	1·0		−1·1		−0·995	1
2·2	1·0		−1·2		−0·980	1
2·3	1·0		−1·3		−0·955	1
2·4	1·0		−1·4		−0·920	1
2·5	1·0		−1·5		−0·875	0·01
2·6	1·0		−1·6		−0·820	1
2·7	1·0		−1·7		−0·755	1
2·8	1·0		−1·8		−0·680	1
2·9	1·0		−1·9		−0·595	1
3·0	1·0		−2·0		−0·500	0·01
3·1	1·0		−2·1		−0·395	1
3·2	1·0		−2·2		−0·280	1
3·3	1·0		−2·3		−0·155	1
3·4	1·0		−2·4		−0·020	1
3·5	1·0		−2·5		0·125	0·01
3·6	1·0		−2·6		0·280	1
3·7	1·0		−2·7		0·445	1
3·8	1·0		−2·8		0·620	1
3·9	1·0		−2·9		0·805	1
4·0	1·0		−3·0		1·000	0·01
4·1	1·0		−3·1		1·205	1
4·2	1·0		−3·2		1·420	1
4·3	1·0		−3·3		1·645	1
4·4	1·0		−3·4		1·880	1
4·5	1·0		−3·5		2·125	0·01
4·6	1·0		−3·6		2·380	1
4·7	1·0		−3·7		2·645	1
4·8	1·0		−3·8		2·920	1
4·9	1·0		−3·9		3·205	1
5·0	1·0		−4·0		3·500	0·01

x	$n = 3$	δ^2	$n = 4$	δ^2_m	$n = 5$	δ^2_m
0·0	1·00000 0	0·030	1·00000 0	0·05999	1·00000 0	0·09996
0·1	0·71483 3	29	0·62933 7	5604	0·54835 4	9020
0·2	0·45866 7	28	0·31473 3	5219	0·18699 7	8094
0·3	0·23050 0	27	0·05233 7	4844	–0·09333 3	7216
0·4	0·02933 3	26	–0·16160 0	4479	–0·30141 9	6385
0·5	–0·14583 3	0·025	–0·33072 9	0·04124	–0·44557 3	0·05601
0·6	–0·29600 0	24	–0·45860 0	3779	–0·53364 8	4860
0·7	–0·42216 7	23	–0·54866 2	3444	–0·57304 6	4163
0·8	–0·52533 3	22	–0·60426 7	3119	–0·57073 1	3492
0·9	–0·60650 0	21	–0·62866 2	2804	–0·53323 3	2899
1·0	–0·66666 7	0·020	–0·62500 0	0·02499	–0·46666 7	0·02330
1·1	-0·70683 3	19	–0·59632 9	2204	–0·37673 3	1799
1·2	–0·72800 0	18	–0·54560 0	1919	–0·26873 6	1308
1·3	–0·73116 7	17	–0·47566 2	1644	–0·14758 7	855
1·4	–0·71733 3	16	–0·38926 7	1379	–0·01781 9	439
1·5	–0·68750 0	0·015	–0·28906 2	0·01124	0·11640 6	0·00059
1·6	–0·64266 7	14	–0·17760 0	879	0·25128 5	–0·00286
1·7	– 0·58383 3	13	–0·05732 9	644	0·38336 6	597
1·8	–0·51200 0	12	0·06940 0	419	0·50953 6	875
1·9	–0·42816 7	11	0·20033 7	0·00204	0·62701 3	1121
2·0	–0·33333 3	0·010	0·33333 3	–0·00001	0·73333 3	–0·01336
2·1	–0·22850 0	9	0·46633 7	196	0·82634 5	1521
2·2	–0·11466 7	8	0·59740 0	381	0·90419 7	1678
2·3	0·00716 7	7	0·72467 1	556	0·96532 5	1806
2·4	0·13600 0	6	0·84640 0	721	1·00844 8	1905
2·5	0·27083 3	0·005	0·96093 7	–0·00876	1·03255 2	–0·01982
2·6	0·41066 7	4	1·06673 3	1021	1·03688 5	2032
2·7	0·55450 0	3	1·16233 7	1156	1·02094 5	2058
2·8	0·70133 3	2	1·24640 0	1281	0·98446 9	2061
2·9	0·85016 7	1	1·31767 1	1396	0·92742 5	2042
3·0	1·00000 0	0·000	1·37500 0	–0·01501	0·85000 0	–0·02002
3·1	1·14983 3	–0·001	1·41733 7	1596	0·75259 1	1942
3·2	1·29866 7	2	1·44373 3	1681	0·63579 7	1863
3·3	1·44550 0	3	1·45333 7	1756	0·50040 5	1766
3·4	1·58933 3	4	1·44540 0	1821	0·34738 1	1652
3·5	1·72916 7	–0·005	1·41927 1	–0·01876	0·17786 5	–0·01523
3·6	1·86400 0	6	1·37440 0	1921	–0·00684 9	1377
3·7	1·99283 3	7	1·31033 7	1956	–0·20531 0	1218
3·8	2·11466 7	8	1·22673 3	1981	–0·41593 1	1047
3·9	2·22850 0	9	1·12333 7	1996	–0·63699 6	863
4·0	2·33333 3	–0·010	1·00000 0	–0·02001	–0·86666 7	–0·00668
4·1	2·42816 7	11	0·85667 1	1996	–1·10299 7	462
4·2	2·51200 0	12	0·69340 0	1981	–1·34393 6	249
4·3	2·58383 3	13	0·51033 7	1956	–1·58735 0	–0·00027
4·4	2·64266 7	14	0·30773 3	1921	–1·83101 9	0·00202
4·5	2·68750 0	–0·015	0·08593 7	–0·01876	–2·07265 7	0·00437
4·6	2·71733 3	16	–0·15460 0	1821	–2·30991 6	677
4·7	2·73116 7	17	–0·41333 0	1756	–2·54039 7	921
4·8	2·72800 0	18	–0·68960 0	1681	–2·76166 5	1168
4·9	2·70683 3	19	–0·98266 2	1596	–2·97124 9	1417
5·0	2·66666 7	–0·020	–1·29166 7	–0·01501	–3·16666 7	0·01665

x	$n = 6$	δ^2_m	$n = 7$	δ^2_m	$n = 8$	δ^2_m
0·0	1·00000 0	0·14984	1·00000 0	0·20307	1·00000 0	0·27929
0·1	0·47172 9	·13060	0·39931 1	·17638	0·33095 4	·22676
0·2	0·07431 7	·11278	—0·02438 9	·14641	—0·11014 7	·18068
0·3	—0·21005 8	9635	—0·30110 6	·11954	—0·36948 1	·14052
0·4	—0·39784 0	8124	—0·45775 3	9556	—0·48728 9	·10582
0·5	—0·50414 5	0·06740	—0·51833 9	0·07430	—0·49836 3	0·07608
0·6	—0·54282 3	5478	—0·50416 0	5557	—0·43251 8	5089
0·7	—0·52651 1	4331	—0·43397 2	3921	—0·31502 8	2980
0·8	—0·46668 7	3294	—0·32417 0	2504	—0·16705 2	0·01244
0·9	—0·37372 4	2363	—0·18895 1	1291	—0·00601 8	—0·00158
1·0	—0·25694 4	0·01532	—0·04047 6	0·00266	0·15399 3	—0·01259
1·1	—0·12466 9	796	+0·11097 7	—0·00586	0·30190 8	2093
1·2	0·01573 1	0·00150	0·25686 4	1279	0·42932 5	2692
1·3	0·15778 7	—0·00411	0·39023 2	1826	0·53020 9	3082
1·4	0·29587 9	892	0·50558 7	2241	0·60060 4	3293
1·5	0·42519 5	—0·01296	0·59875 8	—0·02535	0·63835 9	—0·03348
1·6	0·54168 0	1630	0·66677 8	2722	0·64288 2	3270
1·7	0·64199 2	1896	0·70775 8	2811	0·61490 6	3083
1·8	0·72345 5	2100	0·72078 3	2815	0·55626 9	2805
1·9	0·78402 1	2246	0·70579 4	2743	0·46971 9	2454
2·0	0·82222 2	—0·02338	0·66349 2	—0·02606	0·35873 0	—0·02048
2·1	0·83713 2	2380	0·59523 7	2411	0·22733 0	1603
2·2	0·82832 3	2376	0·50295 7	2169	0·07994 9	1131
2·3	0·79582 9	2330	0·38906 0	1887	—0·07871 7	646
2·4	0·74010 8	2245	0·25635 2	1573	—0·24384 1	—0·00159
2·5	0·66200 1	—0·02126	0·10795 7	—0·01235	—0·41056 9	+0·00321
2·6	0·56269 5	1975	—0·05275 6	878	—0·57412 7	781
2·7	0·44369 2	1796	—0·22223 4	511	—0·72991 5	1219
2·8	0·30677 4	1593	—0·39681 7	—0·00139	—0·87357 2	1624
2·9	0·15396 4	1369	—0·57278 8	0·00234	—1·00105 9	1992
3·0	—0·01250 0	—0·01127	—0·74642 9	0·00601	—1·10870 6	0·02317
3·1	—0·19020 2	869	—0·91408 0	958	—1·19326 8	2597
3·2	—0·37657 5	600	—1·07217 5	1301	—1·25195 2	2825
3·3	—0·56892 9	321	—1·21729 3	1626	—1·28247 6	3005
3·4	—0·76448 5	—0·00036	—1·34619 3	1929	—1·28305 3	3129
3·5	—0·96039 5	0·00252	—1·45584 9	0·02206	—1·25244 0	0·03201
3·6	—1·15378 0	543	—1·54349 1	2457	—1·18992 0	3216
3·7	—1·34174 1	831	—1·60662 0	2677	—1·09533 3	3182
3·8	—1·52139 6	1116	—1·64303 7	2865	—0·96902 9	3090
3·9	—1·68989 6	1397	—1·65086 4	3018	—0·81191 7	2950
4·0	—1·84444 6	0·01668	—1·62857 3	0·03137	—0·62539 8	0·02761
4·1	—1·98233 7	1929	—1·57497 8	3220	—0·41136 0	2522
4·2	—2·10096 1	2179	—1·48926 0	3261	—0·17217 2	2246
4·3	—2·19782 2	2414	—1·37099 6	3270	0·08940 0	1919
4·4	—2·27056 8	2636	—1·22010 7	3235	0·37010	1567
4·5	—2·31699 2	0·02837	—1·03693 1	0·03166	0·66640	0·01171
4·6	—2·33507 7	3022	—0·82216 8	3058	0·97437	756
4·7	—2·32297 7	3186	—0·57690 1	2908	1·28984	0·00304
4·8	—2·27905 6	3329	—0·30261 5	2729	1·60833	—0·00152
4·9	—2·20188 9	3449	—0·00111 2	2508	1·92525	639
5·0	—2·09028 1	0·03544	0·32540 1	0·02248	2·23573	—0·01156

x	$n = 9$	δ_m	$n = 10$	δ^2_m
0·0	1·00000 0	0·35869	1·00000 0	0·44788
0·1	0·26651 5	·28095	0·20585 4	·33823
0·2	−0·18392 9	·21454	−0·24665 4	·24712
0·3	−0·41794 3	·15838	−0·44902 1	·17240
0·4	−0·49188 6	·11140	−0·47634 8	·11200
0·5	−0·45291 9	0·07260	−0·38937 4	0·06407
0·6	−0·34000 1	4110	−0·23633 6	0·02694
0·7	−0·18479 8	0·01601	−0·05465 5	−0·00096
0·8	−0·01254 1	−0·00343	0·12752 1	2099
0·9	0·15719 0	1796	0·28993 1	3441
1·0	0·30974 4	−0·02824	0·41894 6	−0·04235
1·1	0·43472 4	3489	0·50643 9	4581
1·2	0·52538 0	3846	0·54878 4	4569
1·3	0·57804 7	3947	0·54595 6	4278
1·4	0·59163 6	3837	0·50073 6	3777
1·5	0·56716 6	−0·03559	0·41801 8	−0·03128
1·6	0·50735 1	3150	0·30419 8	2383
1·7	0·41622 0	2643	0·16664 6	1586
1·8	0·29878 0	2070	0·01326 0	−0·00777
1·9	0·16071 2	1456	−0·14793 0	0·00014
2·0	0·00811 3	−0·00825	−0·30906 6	0·00759
2·1	−0·15274 3	−0·00197	−0·46273 3	1439
2·2	−0·31560 2	0·00412	−0·60216 5	2036
2·3	−0·47440 3	985	−0·72140 9	2539
2·4	−0·62343 4	1514	−0·81544 3	2942
2·5	−0·75743 3	0·01984	−0·88025 2	0·03236
2·6	−0·87171 2	2390	−0·91289 5	3425
2·7	−0·96221 7	2728	−0·91148 6	3503
2·8	−1·02557 9	2989	−0·87522 7	3483
2·9	−1·05918 4	3177	−0·80432 8	3356
3·0	−1·06116 0	0·03285	−0·70002 6	0·03145
3·1	−1·03042 2	3320	−0·56443 1	2846
3·2	−0·96662 4	3278	−0·40050 6	2475
3·3	−0·87017 8	3162	−0·21195 0	2037
3·4	−0·74223 0	2983	−0·00312 1	1548
3·5	−0·58457 3	0·02735	0·22110 6	0·01014
3·6	−0·39966 8	2431	0·45542 6	0·00455
3·7	−0·19054 9	2073	0·69425 0	−0·00131
3·8	0·03922	1672	0·93175	724
3·9	0·28563	1230	1·16203	1304
4·0	0·54427	0·00748	1·37927	−0·01885
4·1	0·81036	0·00252	1·57766	2435
4·2	1·07893	−0·00267	1·75175	2950
4·3	1·34480	797	1·89639	3427
4·4	1·60269	1334	2·00682	3856
4·5	1·84726	−0·0186	2·07878	−0·0423
4·6	2·07329	238	2·10855	453
4·7	2·27558	287	2·09318	476
4·8	2·44918	335	2·03030	493
4·9	2·58936	379	1·91826	501
5·0	2·69173	−0·0418	1·75628	−0·0500

1.374.4 : 681.142 : 518.5

Monograph No. 137 M
June 1955

FUNCTION GENERATORS BASED ON LINEAR INTERPOLATION WITH APPLICATIONS TO ANALOGUE COMPUTING

By E. G. C. BURT, B.Sc., Associate Member, and O. H. LANGE, Dr. rer. nat.

The paper was first received 18*th November*, 1954, *and in revised form* 16*th March*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in June*, 1955.)

SUMMARY

The use of function generators in electronic analogue computing d simulation greatly extends the range of problems which can be lved by these methods. This paper presents a technique in which ode units are used to approximate to the functions by linear interlation. It is shown that the method can be extended to deal with wide class of functions, including multi-variate functions.

Analogue multiplication and division are discussed as particular cases function generators, and formulae for the general function are eveloped.

The results are presented of an experimental generator for $\sin x$ in e range $-\pi \leqslant x \leqslant \pi$, in which the error is about $1\frac{1}{2}\%$ of the maxium output.

LIST OF SYMBOLS

v = Instantaneous input voltage to function generator.
v_0 = Instantaneous output voltage of function generator.
$_x, v_y$ = Voltages proportional to variables x and y.
V_B = Bias voltage of diode units.
V = Constant input voltage.
i = Current through diodes.
c = Dimensional scale factor.

(1) INTRODUCTION

The role of automatic computing machines—both of the digital d the continuous-variable type—in the analysis and solution of wide variety of problems is becoming increasingly important. hey are rightly regarded as indispensable aids to analytical search, and the growing complexity of the tasks undertaken is sulting in a demand for higher standards of accuracy and exibility from these machines.

The basis of the analogue machine is the provision of a physical stem, the internal state of which is at any instant described by e set of equations whose solution is required, or which obeys e same laws as the system under investigation. Accordingly, e accuracy of the solution will depend on the extent to which exact analogy is obtained, and this in turn depends on the ccuracy of the machine components. The computer will ways provide the exact solution of the equations describing e analogue system, but they may not be the desired equations. Even with perfect components, the accuracy of the analogue ethod is still limited, since the solution must be obtained by easuring the magnitudes of physical quantities—the analogue riables. In a digital machine this is not necessary, except in e gross sense of distinguishing between orders of magnitude, d accuracy is limited only by the amount of equipment required d the time taken to reach a solution. There exists, however, a ide range of problems in which the accuracy furnished by an alogue machine (say 1 or 2%) is sufficient, and where an act computation is neither necessary nor justified; at a later age in the analysis it may be desirable to examine a narrow region in more detail, for which the higher resolution of a digital machine may be required.

The electronic type of analogue machine has been used extensively, because of its intrinsically high operating speed, its ease of construction and its comparatively low cost. The analogue variable is usually a voltage, with or without a carrier; where integration and similar operations are necessary (as in the solution of differential equations) it is more convenient to deal with the direct voltage throughout rather than with a carrier system. The principal source of inaccuracy in the direct method is the zero drift of amplifiers, and such errors can be almost eliminated by automatic stabilization techniques.[1]

The scope of the electronic analogue machine can be greatly extended by the use of function generators, i.e. units in which the output voltage is a specified function of the input voltage or voltages;* the function may be known analytically or simply as tabulated data. Such generators are essential for the solution of non-linear and variable-coefficient equations, for co-ordinate transformations, and for introducing coefficients which have been derived experimentally.

Several types of generators have been described, including those which use shaped potentiometers and photo-electric methods.[2] In the latter system the light spot of a cathode-ray tube is constrained to follow the profile of a mask placed over the face of the tube, the profile corresponding to the function to be generated. Other methods are based on linear interpolation, and those in which diode circuits are used to provide the linear interval offer some advantages: they are inherently fast in operation, and do not require special apparatus. The circuits which have been described in the literature[2,3] are restricted to monotonically increasing functions whose first derivatives are also monotonic and in the same sense, and attention has been focused on multiplication. The multiplier is a particular case of a function generator, since by the use of some such identity as

$$4xy \equiv (x + y)^2 - (x - y)^2$$

the operation is reduced to the generation of the monotonic function z^2, together with adding and subtracting facilities.

It is the purpose of this paper to show that, by suitable combinations of diode circuits and high-gain feedback amplifiers, it is possible to remove the monotonic restriction, so that the technique is available for a wider class of functions, including multivariate functions.

(2) GENERATION OF MONOTONICALLY INCREASING FUNCTIONS

(2.1) Linear Interpolation

The basis of the diode technique is to approximate to the given function by means of straight lines, the number of lines depending on the accuracy required. There is no restriction on

Correspondence on Monographs is invited for consideration with a view to blication.

Mr. Burt and Dr. Lange are at the Royal Aircraft Establishment, Farnborough.

* This definition of a function generator includes such devices as summing amplifiers, integrators, etc., which are commonly employed in analogue machines. In the paper, however, we are concerned with a wider class of functions (e.g. trigonometric functions, with time-independent arguments), the generation of which requires additional apparatus.

the length of each interval, and it is often preferable to use unequal intervals. The diode circuits are then so arranged that the current flowing is proportional to the input voltage throughout a given interval, the current/voltage slope being chosen to correspond to the slope of the straight-line approximation for that interval.

(2.2) The Basic Diode Circuit

A convenient starting-point for the generation of functions is the diode circuit of Fig. 1. This arrangement lends itself to the formation not only of monotonic functions, but also of the wider class of functions described later.

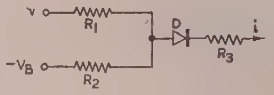


Fig. 1.—Basic diode circuit.

Using the notation of Fig. 1, the current resulting from an input voltage, v, is given by

$$\begin{cases} i = \dfrac{(vG_1 - V_BG_2)G_{3D}}{G_1 + G_2 + G_{3D}}, & v > \dfrac{V_BG_2}{G_1} \\ \quad = 0, & v < \dfrac{V_BG_2}{G_1} \end{cases}$$

where $-V_B$ is a fixed negative voltage, G_1 and G_2 are the conductances of R_1 and R_2, and G_{3D} is the combined conductance of the forward diode resistance and the resistor R_3.

The graph of the current as a function of applied voltage therefore consists of the v-axis for $v < V_BG_2/G_1$, and of a straight line of slope

$$\frac{di}{dv} = \frac{G_1G_{3D}}{G_1 + G_2 + G_{3D}}$$

for $v > V_BG_2/G_1$.

The intercept and the slope can thus be controlled independently by adjustment of the bias $-V_B$ and the resistances R_1, R_2 and R_3.

These equations assume that the diode behaves as a perfect non-conductor for a negative applied voltage, and as a constant resistance for a positive voltage. The actual diode characteristic is modified by the effect of contact potential, but it is shown later (Section 6.1) that the departure from the ideal curve need not greatly affect the accuracy of the unit.

(2.3) Conversion of Current Output to Voltage Output

The addition of a feedback amplifier (Fig. 2) to the circuit of Fig. 1 provides a voltage output v_0 proportional to the diode current. If the internal gain of the amplifier is very high,

$$\left.\begin{aligned} v_0 &= -iR = -\frac{(vG_1 - V_BG_2)G_{3D}R}{G_1 + G_2 + G_{3D}}, \quad v > V_BG_2/G_1 \\ &= 0 \qquad\qquad\qquad\qquad\qquad\quad v < V_BG_2/G_1 \end{aligned}\right\} \quad . \quad (1)$$

where R is the feedback resistance.

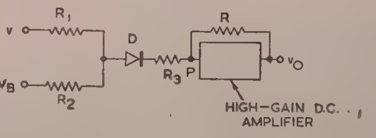


Fig. 2.—Conversion to voltage output.

The high gain ensures that the output voltage v_0 is avail at a low impedance for subsequent stages, and also that point P remains very nearly at ground potential—its excur can be kept as small as desired by making the internal sufficiently high.

(2.4) Multiple Diode Units

In order to build up the straight lines which approximat the desired function, one diode unit is required for each line. Since the amplifier input grid is effectively at gro potential, it follows that the outputs of the diode units ca connected together at this point; the contribution from diode will be unaffected by the currents from the remai units, and the output voltage will be proportional to the su the currents. The arrangement is in fact a summing ampl the number of inputs being controlled by the input voltage it

(2.5) Parabola

As a simple example of the function generator, consider approximation to the parabola f(v) = cv^2 by linear interpola between the points whose abscissae are 0, a_1, a_2, a_3, etc. (Fig

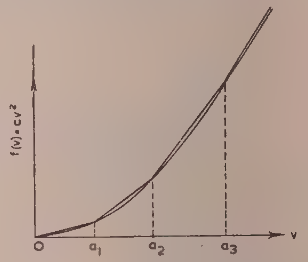


Fig. 3.—Linear interpolation for the parabola.

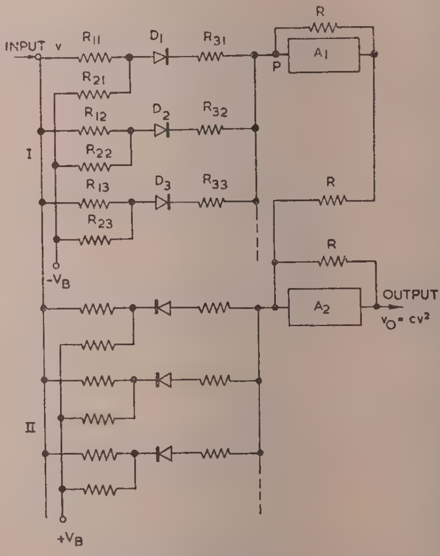


Fig. 4.—Circuit for the parabola.

The circuit is shown in Fig. 4. The slope of the line j the points (a_r, ca_r^2) and (a_{r+1}, ca_{r+1}^2) is $c(a_r + a_{r+1})$, so th first diode unit D_1 is required to supply a current such tha

$$-\frac{dv_0}{dv} = ca_1 \text{ for } a_1 \geqslant v \geqslant 0$$

he negative sign arising from the reversing property of the mplifier.

Inserting these values in eqn. (1),

$$\frac{G_{11}G_{31}R}{G_{11}+G_{21}+G_{31}} = ca_1$$

nd $$V_BG_{21}/G_{11} = 0$$

re the necessary conditions.

For convenience the diode resistance is omitted from these and ubsequent equations: it can be regarded as forming part of the esistance R_{31}, etc.

In this case R_{21} is infinite, since the diode must start conducting when $v = 0$ and therefore requires no bias.

The second diode unit D_2 is required to provide current at the point (a_1, ca_1^2) (Fig. 3) to accommodate the increased slope beyond this point. The first diode continues to conduct, so that D_2 must account for the differential slope $c(a_1 + a_2) - ca_1 = ca_2$.

Thus $$G_{12}G_{32}R/(G_{12}+G_{22}+G_{32}) = ca_2$$

and $$V_BG_{22}/G_{12} = a_1$$

The diode D_3 must be biased at $v = a_2$ and provide a slope $c(a_3 + a_2) - c(a_2 + a_1) = c(a_3 - a_1)$, and so on. The conditions for the rth diode are

$$\frac{G_{1r}G_{3r}}{G_{1r}+G_{2r}+G_{3r}} = c(a_r - a_{r-2}) \quad . \quad . \quad . \quad (2)$$

and $$V_BG_{2r}/G_{1r} = a_{r-1}$$

The output of the amplifier A_1 (Fig. 4) is equal to $-cv^2$ for $v \geqslant 0$, if the resistances satisfy the conditions of eqns. (2). For $v < 0$, the output is zero, since all the diodes of group 1 will be biased negatively. In order to obtain the correct output for negative inputs, a second group of diode units (group 2 in Fig. 4) is required, in which the diodes and the biasing voltage are reversed. This group functions in the same way as group 1, except that the output of A_2 is equal to $+cv^2$ when $v \leqslant 0$, and zero when $v > 0$.

If now the output of A_1 is connected via a resistor R to the input of A_2, the output of the latter will be cv^2 for positive and negative values of v, as required.

(3) MULTIPLICATION

For a voltage-analogue machine in which a multiplier forms an intermediate element, it is necessary to express as a voltage each of the variables whose product is required, and the output must also be a voltage. The dimension (volts)^2 does not occur physically, so that either the multiplication must be indirect or the input voltages must first be transformed into quantities whose product forms a third physical quantity, as in the relation $V = IR$. Such conversions inevitably involve a time delay, as do various indirect methods[4] in which modulation forms the basis of the multiplication. The diode method has a much faster response, the limiting factor being the effect of diode capacitances and the time-constants of the associated amplifiers.

By means of the identity

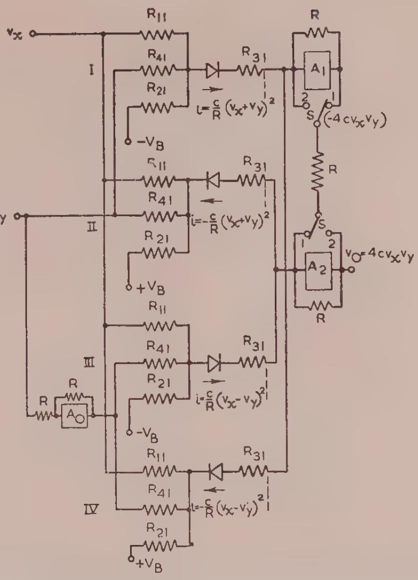
$$4cv_xv_y = c[(v_x + v_y)^2 - (v_x - v_y)^2]$$

the squaring circuits of Section 2.5 can be used to give the function

$$v = 4cv_xv_y$$

where v_x, v_y are proportional to x and y, the variables whose product is required, and c is a dimensional scale factor.

The arrangement is shown in Fig. 5. Each of the four groups of diodes, of which only the first diode unit is shown, is similar to the groups of Fig. 4, except for the additional resistors R_{4r} to which the v_y input is connected. Groups 1 and 2 form the parabola $(v_x + v_y)^2$, and groups 3 and 4 provide $(v_x - v_y)^2$. Groups 1 and 3 conduct for positive values of $(v_x + v_y)$ and $(v_x - v_y)$ respectively, while groups 2 and 4 deal with negative values.



v_x
v_y
R_{11}
R_{41}
R_{21}
R_{31}
R
I
II
III
IV
$-V_B$
$+V_B$
A_1
A_2
A_0
S
1
2
$i = \frac{c}{R}(v_x+v_y)^2$
$i = -\frac{c}{R}(v_x+v_y)^2$
$i = \frac{c}{R}(v_x-v_y)^2$
$i = -\frac{c}{R}(v_x-v_y)^2$
$(-4cv_xv_y)$
$v_o = 4cv_xv_y$

Fig. 5.—Diode multiplier.

The outputs of the diode groups are connected to the amplifiers A_1 and A_2 in such a way that the output of A_2 gives the correctly-signed product v_xv_y under all conditions. The action of the circuit is best explained as follows:

		Output of A_1	Output of A_2
$v_x + v_y > 0$ $v_x - v_y > 0$	Groups 1 and 3 only conduct	$-c(v_x + v_y)^2$	$4cv_xv_y$
$v_x + v_y > 0$ $v_x - v_y < 0$	Groups 1 and 4 only conduct	$-4cv_xv_y$	$4cv_xv_y$
$v_x + v_y < 0$ $v_x - v_y < 0$	Groups 2 and 4 only conduct	$c(v_x - v_y)^2$	$4cv_xv_y$
$v_x + v_y < 0$ $v_x - v_y > 0$	Groups 2 and 3 only conduct	0	$4cv_xv_y$

Thus the output of A_2 is the correct product $4cv_xv_y$, with v_x and v_y unrestricted and unrelated in sign.

If the switch S (Fig. 5) is thrown to the second position, the output of A_1 is then $-4cv_xv_y$. The reversing amplifier A_0 is an essential part of the system, but it obviates the need for an external reversing amplifier for cascading with subsequent stages, since either the positive or negative product can be selected.

(4) GENERAL FUNCTIONS

(4.1) General Functions with Positive Arguments

The applications discussed so far have been confined to monotonically increasing functions whose first derivatives are also monotonic. In order to decrease the slope at a given point, it is necessary to subtract currents from the input to the amplifier. For example, suppose that it is required to produce the simple curve of Fig. 6(d) for positive values of v. Since the first straight

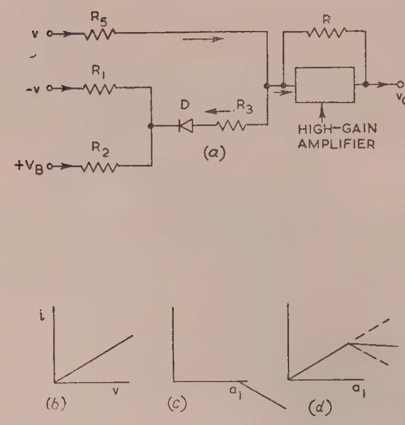


Fig. 6.—Circuit for decreasing slopes.
(a) Diode unit.
(b) Current due to direct input.
(c) Current due to diode circuit.
(d) Combined current.

line starts from zero, it is obtained merely by applying the input v to the amplifier [Fig. 6(a)] through an appropriate resistor R_5. If a voltage $-v$ is applied to the diode unit, its current contribution will be as shown in Fig. 6(c), where a_1 is given by

$$a_1 = V_B G_2 / G_1, \text{ as before.}$$

The output v_0 will be proportional to the sum of the currents of Figs. 6(b) and 6(c), i.e. the desired function of Fig. 6(d).

A combination of the normal diode unit of Fig. 1 and the reversed circuit of Fig. 6(a) provides the means for approximating to a general function, provided that the latter is single-valued.

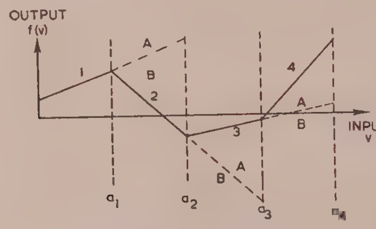


Fig. 7.—Construction of a general function for a positive argument.

Fig. 7 illustrates the construction of a general function for positive values of the argument. The initial interval 1 may or may not require a diode, depending on whether the curve starts from the v-axis or the f(v) axis. In the case shown, the first interval is obtained by providing a constant input proportional to f(0) together with a direct v input to give the correct slope. If it is desired to alter the slope at a_1, a diode unit must be added, its mode of connection depending on whether the existing slope is to be increased (region A) or decreased (region B). The regions A and B are bounded by the slope of the previous interval and the verticals at the change-over points. In the example of Fig. 7 the second interval is in region B, so that a reversed diode with input $-v$ and bias $+V_B$ is necessary. At a_2 the slope increases (region A), requiring a normal diode unit with input $+v$ and bias $-V_B$. Theoretically this process can be continued for as many stages as desired, all the current contributions being added by the same amplifier; in practice, however, the accuracy diminishes as the range is extended, since the result then depends on the addition and subtraction of a large number of currents.

The reversal in sign of the function f(v) (e.g. Fig. 7) causes r difficulty; the cross-over points are not significant in the dioc technique, since no modification of the circuit is called for.

(4.2) Extension to General Function with Negative Argumen

For negative values of the argument v, all the diodes in tl circuits of Section 4.1 are non-conducting, and there will be r output apart from that due to any direct inputs. If the functic f(v) is required for $v < 0$, further diodes are required with reverse connections and reversed bias voltages: the operation of suc circuits has already been described for the parabola (Section 2.5 Thus, for a function having a region of decreasing slope fe $v > 0$, and a similar region for $v < 0$, two groups of reverse diodes are required—one for each region.

(4.3) Formulae for General-Function Generator

The necessary formulae and modes of connection for a gen rator representing a general function can now be stated.

Let the points between which linear interpolation is requir be (a_0, h_0), (a_1, h_1) . . . for $v \geqslant 0$, and (b_0, k_0), (b_1, k_1) . . . f

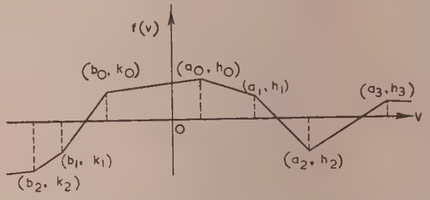


Fig. 8.—Interpolation for the general function.

$v < 0$ (Fig. 8). These points are not necessarily on the cur f(v). The slope of the line joining the points (a_{r-1}, h_{r-1}) ar (a_r, h_r) is

$$\frac{h_r - h_{r-1}}{a_r - a_{r-1}}$$

and the slope of the previous interval is

$$\frac{h_{r-1} - h_{r-2}}{a_{r-1} - a_{r-2}}$$

The diode unit which is biased at $v = a_{r-1}$ must provide t difference in these slopes, so that

$$\frac{G_{1r}G_{3r}R}{G_{1r} + G_{2r} + G_{3r}} = \left|\frac{h_r - h_{r-1}}{a_r - a_{r-1}} - \frac{h_{r-1} - h_{r-2}}{a_{r-1} - a_{r-2}}\right| \quad .$$

and
$$|V_B|\frac{G_{2r}}{G_{1r}} = a_{r-1} \quad . \quad . \quad . \quad . \quad .$$

These equations define the resistances associated with the r diode for $r = 1, 2, 3, \ldots$ (for $r = 1, a_{-1} = b_0$).

Similarly the rth diode, biased at $v = b_{r-1}$, must satisfy t equations

$$\frac{G_{1r}G_{3r}R}{G_{1r} + G_{2r} + G_{3r}} = \left|\frac{k_r - k_{r-1}}{b_r - b_{r-1}} - \frac{k_{r-1} - k_{r-2}}{b_{r-1} - b_{r-2}}\right| \quad .$$

and
$$|V_B|\frac{G_{2r}}{G_{1r}} = - b_{r-1} \quad . \quad . \quad . \quad . \quad .$$

(For $r = 1, b_{-1} = a_0$.)

The interval between b_0 and a_0 requires an input v (or $-v$) a resistance R_5 [Fig. 6(a)], where

$$\frac{R}{R_5} = \left|\frac{h_0 - k_0}{a_0 - b_0}\right| \quad . \quad . \quad . \quad . \quad . \quad .$$

and a constant input V via R_6 such that

$$V\frac{R}{R_6} = -\,\mathrm{f}(0) \text{ if the output is } \mathrm{f}(v)$$

or

$$+\,\mathrm{f}(0) \text{ if the output is } -\mathrm{f}(v) \quad . \quad . \quad (8)$$

For an output f(v), the input is $-v$ if the slope of the interval between b_0 and a_0 is positive, and $+v$ if the slope is negative. If the output $-$f(v) is required, the sign of v is reversed.

Eqns. (3)–(8) define the necessary relations between all the resistances of the complete function generator. Each diode unit can be connected in eight different ways, as shown in Fig. 9.

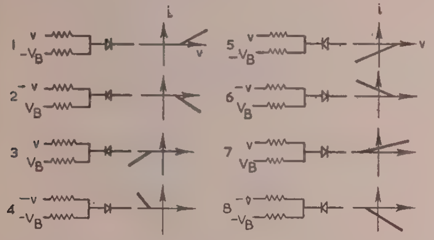


Fig. 9.—Modes of connection for the basic diode unit.

The relevant modes for the above equations are modes 1 to 4 of Fig. 9, and the correct mode for each diode unit can be found from Table 1, in which

$$\Delta(a_r) = \frac{h_r - h_{r-1}}{a_r - a_{r-1}} - \frac{h_{r-1} - h_{r-2}}{a_{r-1} - a_{r-2}}$$

and

$$\Delta(b_r) = \frac{k_{r-1} - k_{r-2}}{b_{r-1} - b_{r-2}} - \frac{k_r - k_{r-1}}{b_r - b_{r-1}}$$

Table 1

Mode of Diode Unit biased at $v = a_{r-1}$, or b_{r-1}

$\Delta(a_r)$	$\Delta(b_r)$	For output $f(v)$	For output $-f(v)$
		Mode (as defined in Fig. 9)	
Positive		2	1
Negative		1	2
	Negative	4	3
	Positive	3	4

Eqns. (3)–(8) and Table 1 specify completely a diode generator for the general function f(v). Two high-gain amplifiers are required—a reversing amplifier to provide $-v$ and the summing amplifier.

By a suitable switching arrangement it is possible to obtain either f(v) or $-$f(v) as the output, the modes being switched according to Table 1.

Modes 5–8 of Fig. 9 have not been used in the above discussion, since the first four are sufficient to make up the general function. In certain cases it may be advantageous to use these additional modes (see, for example, Section 7.2), but eqns. (3)–(8) are then slightly modified.

(4.4) Generators for Sine and Cosine Functions

(4.4.1) Sine Generator.

As an example of the general function, the arrangement for f(v) = $V \sin cv$ is shown in Fig. 10. The switch S is shown in position 1, and the output is then $-V \sin cv$. For the interval 0 to π the slope is decreasing [$\Delta(a_r)$ negative], so that, from Table 1, the diodes for this region are connected as in mode 2. The number of diodes in this group (group 1) depends on the number of intervals chosen to interpolate sufficiently accurately to sin cv over the interval 0 to π, but only two diode units of each group are shown in the Figure.

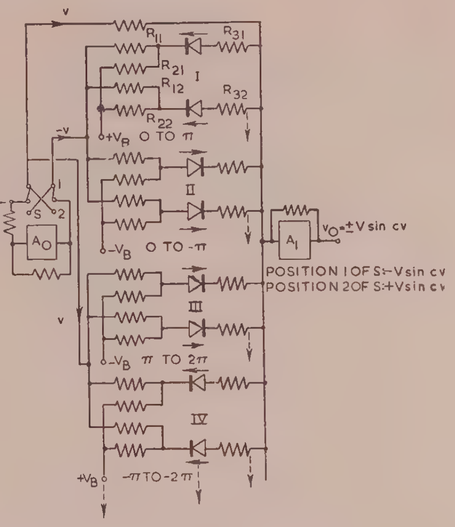


Fig. 10.—Generator for f(v) = ± $V \sin cv$.

From π to 2π, $\Delta(a_r)$ is positive, so that the appropriate mode for group 3 is mode 1. The diodes of group 2 deal with the range 0 to $-\pi$, in which $\Delta(b_r)$ is positive, so that mode 4 is required. Finally, group 4 has mode 3, since in the interval $-\pi$ to -2π, $\Delta(b_r)$ is negative.

The range of v can be extended by adding further groups of diodes, their modes of connection being determined in exactly the same way. The accuracy however diminishes as the range is extended, as noted in Section 4.1.

In Fig. 10, use is made of the symmetry of the sine function to simplify the switching for obtaining $\pm V \sin cv$. For example, if the switch is in position 2 (output $V \sin cv$) the connections of group 1 change from mode 2 to 3, which is the correct mode for the interval 0 to $-\pi$ when the output is +f(v).

(4.4.2) Cosine Generator.

Fig. 11 illustrates the arrangement for f(v) = ± $V \cos cv$. Here f(0) $\neq$ 0, so that from eqn. (8) a constant input is required in addition to the diode stages. The latter are similar to those of the sine generator, except that $\Delta(a_r)$ and $\Delta(b_r)$ change sign at $\pm\pi/2$, $\pm 3\pi/2$, etc., instead of $\pm\pi$, $\pm 2\pi$, so that the groups are 0 to $\pm\pi/2$, $\pm\pi/2$ to $\pm 3\pi/2$, etc. Again, either plus or minus $V \cos cv$ can be obtained through the switch S.

(4.5) Multivariate Functions

With the aid of the multiplier of Section 3.2 and the general function generators of Section 4.3, it is possible to generate a multivariate function f($v_x, v_y, v_z, \ldots$), where $v_x, v_y, \ldots$ are independent variables, provided that the function is single-valued, finite and continuous in the region required. The function can also contain tabulated functions provided that each is a function of one variable only.

As a simple example, consider the function

$$\mathrm{f}(v_x, v_y, v_z) = \arctan\left[\frac{|(v_x^2 + v_y^2)^{\frac{1}{2}}|}{v_z}\right].$$

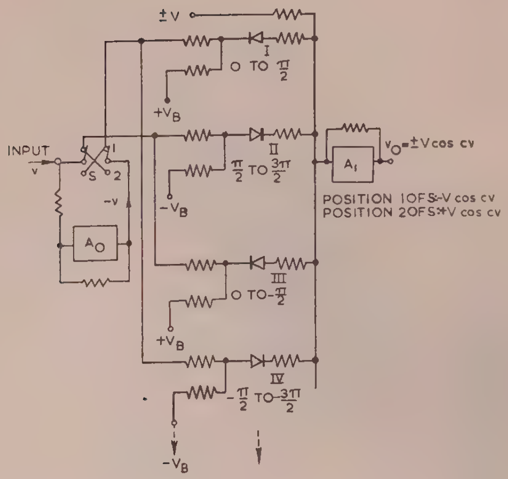


Fig. 11.—Generator for $f(v) = \pm V \cos cv$.

This would require two function generators of the type v^2, and one each of $|v^{\frac{1}{2}}|$, $1/v$ and arc tan v, together with a multiplier.

(5) DIVISION

If a multiplier is available, division can be accomplished in two ways—by first generating the reciprocal of the divisor and then multiplying, or by using the multiplier in a feedback circuit. The application of the diode method to each of the systems is discussed briefly below.

(5.1) The Reciprocal Generator

Since the function $f(v) = 1/v$ has a singularity at the origin, it is necessary to limit the curve in this region. If v is not required to change sign, it is sufficient to limit $1/v$ to a constant maximum value for small values of v. If the divisor does change sign, it is necessary to join the positive and negative limiting values by a steep line—the slope depending on the accuracy required in the vicinity of $v = 0$. The unit can be designed for this curve from the formulae of Section 4.3; the function $f(v_x, v_y) = v_y/v_x$ is then obtained by applying v_y and $1/v_x$ to a multiplier—such as that of Section 3.2.

The correct sign of v_y/v_x will be preserved for either sign of v_x or v_y, but the magnitude of the output will be in error for small values of v_x within the region of the approximation.

(5.2) Multiplier used as a Divider

Fig. 12 indicates how the diode multiplier, with the addition of the amplifier A_3, may be used to give

$$v_0 = v_y/v_x$$

The inputs to the multiplier are v_0 and v_x, giving the negative product. This is added to v_y at the input of A_3, which has a gain α. The output of A_3 is therefore $\alpha(v_y - v_x v_0)$, which provides the input to the multiplier.

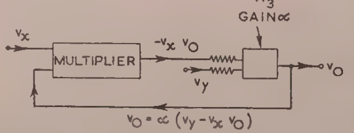


Fig. 12.—Diode multiplier used as a divider.

Thus

$$\alpha(v_y - v_x v_0) = v_0$$

so that for $\alpha \rightarrow \infty$

$$v_0 = v_y/v_x$$

provided that the usual stability conditions are satisfied.

Both $+v_0$ and $-v_0$ are required as inputs to the multiplier, so that the unit provides plus and minus v_y/v_x simultaneously. In this type of divider the divisor v_x must not be allowed to change sign, since the feedback would then become regenerative instead of degenerative.

Although restricted in the sign of the divisor, the circuit requires only one amplifier in addition to the multiplier, whereas for the reciprocal divider an additional function generator is necessary. Also, the range of the reciprocal generator is rather limited, so that the feedback multiplier is to be preferred where the divisor is not required to change sign.

(6) PRACTICAL CONSIDERATIONS

(6.1) Diode Characteristics

It has been assumed so far that the diode unit behaves as a perfect resistance for positive voltages, and passes no current for negative voltages. Account must be taken of the actual diode characteristic—the tail current due to contact potential, the effect of heater voltage, and variations from diode to diode—so far as they affect the operation and stability of the function generator.

It is clearly desirable to be able to calculate the resistances for a function generator from the equations developed in Section 4, so that the unit can be constructed without reference to particular diodes and without experimental adjustment for each unit. This requirement imposes the necessity of uniformity in the diodes themselves, particularly with regard to contact potential.

Some work has been carried out in this connection,[5] mainly on the CV140 double diode, with the object of achieving an accuracy of within 1% for a generator based on calculated values rather than on experimental curve fitting. To this end it was found necessary to reject about 25% of these valves as having contact potentials outside the admissible range; if the remainder were aged for 200–300 hours at about their full rating, their current/voltage relations remained stable for long periods.

The constancy of the heater voltage was also found to be important, and the variation from diode to diode appeared to be less for heater voltages rather lower than the normal 6·3 volts.

It was found that for a series resistance R in the range of kilohms to megohms, the current/voltage relationship could be represented fairly accurately as

$$i = (v + \Delta)/1{\cdot}005R$$

i.e. the effective diode resistance is 0·5% of the series resistance, and the effect of contact potential is such that current begins to flow for a voltage of $-\Delta$.

This simple relation is useful in that the generator can be designed on the assumption of perfect diodes, if eqns. (4) and (6) of Section 4.3 are replaced by

$$|V_B|G_{2r}/G_{1r} = a_{r-1} + \Delta$$

and

$$|V_B|G_{2r}/G_{1r} = -b_{r-1} + \Delta$$

and if the practical value of R_{3r} is made $\frac{1}{2}$% less than the calculated value, to allow for the diode resistance.

For a given heater voltage, the voltage Δ is a function of the series resistance and the voltage range over which the linear approximation is required; for the CV140, Δ lies between 0 and

0·2 volt. If the accuracy is required near cut-off, the higher value of Δ is used. On the other hand, if the accuracy for large inputs is of more importance, the lower value of Δ is more appropriate.

For more accurate work, the approximation can be improved by taking into account the variation of diode resistance as a unction of the series resistance for different ranges of current output. However, this complicates the calculation considerably, and is only worth while if the resistors themselves are very stable, e.g. wire-wound resistors.

(6.2) Other Factors

The accuracy of the function generator depends not only on he diodes, but also on the stability of the resistors, the bias voltages $\pm V_B$, and the high-gain amplifiers. The voltage source should therefore be stabilized and the amplifiers should be of he relay-corrected type[1] to eliminate drift as a source of error. For very accurate work it is essential to use wire-wound resistors hroughout, but for many applications high-stability carbon resistors suffice, particularly if a generous margin is allowed in heir power dissipation and if the ambient temperature is kept airly constant.

(7) EXPERIMENTAL RESULTS FOR A SINE GENERATOR

Fig. 13 shows the results obtained for an experimental sine-wave generator constructed according to the circuit of Fig. 10. Seven tangents were chosen to approximate the sine curve over he range 0–π, requiring three double diodes (type CV140).

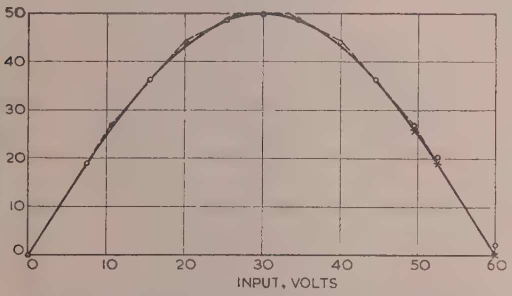


Fig. 13.—Experimental results for the sine function.

○ Experimental points with resistors as calculated.
× Experimental points after adjustment of R_{35} and R_{36}.

The resistances R_{11}–R_{16} (Fig. 14) were chosen to give reason-ble current levels, and the remaining resistances were calculated rom the general formulae of Section 4.3, allowing for the diode haracteristics as in Section 6.1. The resulting circuit is shown n Fig. 14, which represents group 1 of Fig. 10. Group 2 has he same value of resistances, but a different mode of connection, o that for the range of ±π a total of six double diodes were ised.

The scale factors used gave a maximum output of ±50 volts, nd for the v input, 1 volt corresponded to 3°, giving a total input ange of ±60 volts for ±π.

It will be seen from Fig. 13 that the tangent approximation eviates from the sine curve by about 2% of the maximum output. Except for the last three points, the experimental points are within ·5% of the tangents, and the maximum error occurs at the orner points. This error is due to the tail of the diode charac-eristic; if the interpolation is suitably arranged, the rounding-off ffect improves the approximation. Thus in Fig. 13 the maxi-num error from the sine curve is 1·5%, compared with the angent error of 2%, apart from the last three points.

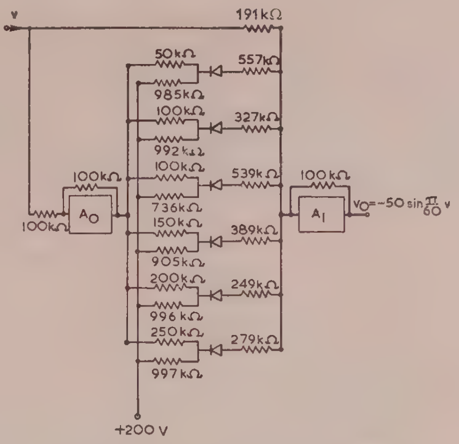


Fig. 14.—Circuit for the sine function.

Resistors: 0·75 watt, high-stability carbon.
Diodes: CV140.

The reason for the latter discrepancy is as follows: for $cv = \pi$ the total current subtracted by all the diodes should equal that provided by the direct input, to give a resultant output of zero. Thus a small error in slope, particularly in the early stages, can give rise to a relatively large error for large values of v, since the output depends on the difference of two large quantities.

It was found that, by adjusting R_{35} and R_{36}, it was possible to correct the slope and bring the last three points (Fig. 13) on to the curve, giving an overall error of about 0·5% (from the tangents) for the range $-\pi$ to π. The error from the sine curve is about $1\frac{1}{2}$%, but this can be improved, within limits, by taking more tangents. However, if the range of cv is extended beyond $\pm\pi$, the adjustment will affect more stages and become more critical. This condition can be mitigated by providing clipping diodes,[5] so that the contribution of a particular stage reaches a constant maximum value instead of rising indefinitely.

Alternatively, it is possible to make use of modes 5–8 (Fig. 9) to obtain the same effect; for example, if the resistor R_5 (Fig. 14) is replaced by a mode-5 diode unit, the latter makes no further contribution beyond its biasing point (in this case it is also necessary to add a constant voltage via a resistor, in order to obtain zero output when $v = 0$). Other combinations are possible, and the choice between them rests on the requirements in a particular case.

(8) CONCLUSIONS

The techniques discussed in the paper appear to offer a convenient means of extending the possibilities of analogue computing and simulation. The diode function generator has a number of advantages, i.e. accuracy, simplicity, flexibility, negligible time delay, and a low output impedance combined with a wide output range.

The results presented for the sine generator show that an accuracy of 1–2% of the maximum output can be achieved without difficulty. Further work is required to establish by how much this figure can be improved.

(9) ACKNOWLEDGMENTS

Acknowledgments are made to the Chief Scientist, Ministry of Supply, and to the Controller of Her Majesty's Stationery Office for permission to publish the paper. The authors are indebted to Mr. K. V. Diprose, with whom the preliminary discussions on the use of diodes took place.

(10) REFERENCES

(1) LANGE, O. H., BURT, E. G. C., and HOLBOURN, C. R.: "A Drift-Compensated D.C. High-gain Amplifier for Summation and Integration," *R.A.E. Technical Note No. G.W.*75.

(2) HARDER, E. L., and CARLETON, J. T.: "New Techniques on the Anacom Electric Analogue Computer," *Transactions of the American I.E.E.*, 1950, **69**, p. 547.

(3) MARSHALL, B. O., JR.: "An Analogue Multiplier," *Nature*, 1951, **167**, p. 29.

(4) THOMAS, W. R., and SQUIRES, M.: "Electronic Analogue Methods of Multiplication," *R.A.E. Technical Note No. G.W.*53, *Aeronautical Research Council*, **13**, p. 170.

(5) LANGE, O. H., and HERRING, G. J.: "Some Electronic Multipliers based on Diode Function Shapers," *R.A.E. Technical Note No. G.W.*245.

21.317.784.029.6 : 621.372.413

Monograph No. 138 R
June 1955

A RESONANT-CAVITY TORQUE-OPERATED WATTMETER FOR MICROWAVE POWER

By R. A. BAILEY, Ph.D., B.Sc., Graduate.

(*The paper was first received* 19*th October*, 1954, *and in revised form* 21*st March*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in June*, 1955.)

SUMMARY

A sensitive method of microwave power measurement is described hich makes use of the mechanical force exerted by the electroagnetic field on a small vane in a resonant cavity. It is shown that e force on the vane is a simple function of the Q-factor of the cavity, e power absorbed in it and the perturbation of its resonant frequency used by the vane. The results of a comparison between an experiental wattmeter based on this principle and a water calorimeter are ven, and the requirements of a practical instrument are discussed.

(1) INTRODUCTION

Most of the devices commonly used for accurate power easurement at microwave frequencies absorb the power in a ssy material and measure the resultant temperature rise. 'alibration is carried out by the application of power from d.c. ources, and the assumption is made that the temperature rise is e same for equal d.c. and microwave powers. The thermistor, olometer, thermocouple, and water calorimeter are all used in is way; none of them gives an absolute measurement of power, lthough the calorimeter is generally regarded as a standard.

More recently the effect of radiation pressure at microwave equencies has been demonstrated by Carrara and Lombardini,[1] nd used by Cullen[2, 3, 4] to measure the power in a waveguide. ı its later form Cullen's instrument consists of a rectangular uide containing a flat metal vane suspended so that it can otate about an axis parallel to the broad dimension of the uide. The electric field of the H_{01} mode exerts a force on the ane which tends to rotate it to the transverse position. The orque on the vane is proportional to the power, and the constant f proportionality may be found from a subsidiary experiment sing a movable piston and standing-wave indicator. The libration is effected by means of measurements of mass, length nd time only, and hence the instrument makes an absolute easurement of power. It is used as a transmission wattmeter nd absorbs negligible power from the guide. With this type of ispension the vane acts, roughly, as a voltmeter across the uide and hence for accurate power measurement the standing-ave ratio of the load must be nearly unity. The magnitude of e torque is a function of the electric field strength and the vane ze. In Cullen's instrument it is of the order of 10^{-4} dyne-n/watt which is rather small for measurements other than boratory ones at powers less than 10 watts.

Evidently, if the power to be measured is fed into a cavity sonator of high Q-factor containing a vane, the field at the ne (and hence the deflection sensitivity) may be made large. urther, it can be shown that the power absorbed in the cavity ay be calculated from the force acting on the vane without any rect knowledge of the field distribution in its vicinity.

An instrument based on this principle has been made and its sign is described.

(2) THEORY

The basis of the calculation for the force acting on the vane is a theorem of adiabatic invariance expressed by Maclean,[5] which states that in a lossless electromagnetic resonator the action of each mode, i.e. the product of total energy and period, is invariant against an adiabatic deformation.

Consider a lossless system consisting of a cavity, resonant in one mode with period τ, containing a small vane and having stored energy W. In general, a force will be exerted on the vane. Let it move slowly a small distance ds. This movement will alter the resonant period of the cavity, but, by the theorem, we have

$$W\tau = \text{a constant}$$

from which

$$Wd\tau + \tau dW = 0$$

and

$$dW = -\frac{W}{\tau}d\tau$$

Since the system is lossless, the change in energy stored by the field must be equal to the work done by the moving vane. Therefore

$$Fds = -dW = W\frac{d\tau}{\tau}$$

$$F = \frac{W}{\tau}\frac{d\tau}{ds}$$

where F is the force acting in the direction ds.

In a practical system the resonator will not be lossless, but the field distribution in a resonator of high Q-factor will be negligibly different from that in a lossless one, and for a given stored energy the force on the vane will be the same.

In a lossy resonator

$$Q = \frac{2\pi \times \text{energy stored}}{\text{Energy lost per cycle}}$$

$$= \frac{2\pi W}{P\tau}$$

where P is the power fed into the cavity. Therefore

$$W = \frac{QP\tau}{2\pi}$$

and

$$F = \frac{QP}{2\pi}\frac{d\tau}{ds}$$

A better form for this expression is obtained if Q is replaced by $\pi f t_0$, where t_0 is the time-constant of decay of oscillation amplitude in the cavity, and f is the frequency. Then

$$F = \frac{\pi f t_0 P}{2\pi}\frac{d\tau}{ds}$$

$$= P\frac{t_0}{2}\frac{1}{\tau}\frac{d\tau}{ds} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Correspondence on Monographs is invited for consideration with a view to blication.

Dr. Bailey is at the Radar Research Establishment, Ministry of Supply.

$$= -P\frac{t_0}{2}\frac{1}{f}\frac{df}{ds} \quad . \quad . \quad . \quad . \quad . \quad (1a)$$

For a rotating vane $\quad T = -P\dfrac{t_0}{2}\dfrac{1}{f}\dfrac{df}{d\theta}$

where T is the torque on the vane.

For microwave cavities of similar shape and mode, Q/λ is inversely proportional to the skin depth δ. Now

$$\delta = \sqrt{\frac{\rho}{\pi\mu f}}$$

where ρ is the resistivity of the cavity wall. Therefore

$$t_0 = \frac{Q}{\pi f} \quad \text{and} \quad \propto \frac{1}{(f)^{3/2}}$$

Therefore the sensitivity of the device as a power meter is inversely proportional to (frequency)$^{3/2}$ for a constant fractional rate of detuning, $\dfrac{1}{f}\dfrac{df}{d\theta}$.

(3) SOME PRACTICAL CONSIDERATIONS

Slater[6] has shown that if a cavity is perturbed, by pushing in its boundary, the change of resonant frequency is dependent on the integral $\int (H^2 - E^2)dv$ over the volume which is removed from the cavity by the perturbation of the wall. The resonant frequency decreases if the perturbation is made at a point of predominant electric field, and increases if the perturbation is made at a point of predominant magnetic field. At some intermediate point it is possible for the effects to cancel. Similarly, for a simple small vane, df_0/ds, and hence the force, is greatest if the vane is placed in a region where one field is predominant and if the vane is of such a shape that it couples more strongly to that field than to the other.

A cylindrical H_{011} resonator was used to test the theory expressed above. At the mid-section along its length the E field in this cavity is circumferential, with a maximum at 0·48r, where r is the radius of the cavity. The H field is axial, having maxima at the axis and the wall and a zero at 0·62r. Vanes were mounted on radial polystyrene rods and the resonant frequency and Q-factor of the cavity were measured as the rods were rotated.

A short, straight copper rod mounted parallel to the axis and placed in the strong E field at about half the radius caused a decrease of resonant frequency as it was rotated from the axial to the transverse position. The change of frequency with rotation roughly followed a $\sin^2\theta$ law. The Q-factor of the cavity decreased as the vane was rotated; hence the force on the vane, which is proportional to the product of Q and $df_0/d\theta$, reached a maximum when the vane made an angle of about 38° with the axis. In Fig. 1, typical curves of Q, Δf_0, and $Qdf_0/d\theta$ are shown.

A wire loop or a disc, suspended so that its centre was on the cavity axis, caused an increase of resonant frequency as its plane was rotated from an axial to a transverse position. By displacing the centre of a vane of this type off the axis it was possible to find a point where its coupling to the magnetic and electric fields was such that its rotation caused very little change in the resonant frequency.

For a given value of $df_0/d\theta$, a straight rod caused less depression of Q than a loop; it was therefore decided to use a rod in the experimental instrument built. The instrument was checked against a water calorimeter.

(4) CONSTRUCTION

A drawing of the cavity and vane system is given in Fig. 2. The cavity used was an S-band echo box, 6 in in diameter and

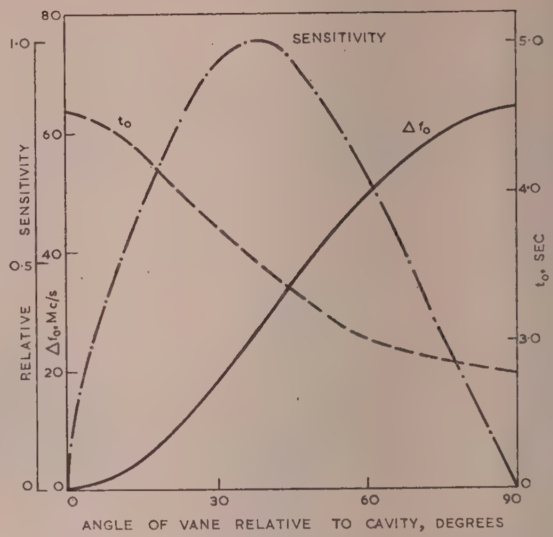


Fig. 1.—Typical curves of t_0, Δf_0, and sensitivity $\left(t_0\dfrac{df_0}{d\theta}\right)$ for a r in an H_{011} cavity.

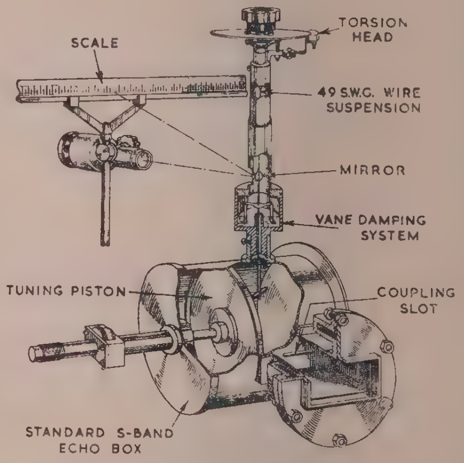


Fig. 2.—Experimental wattmeter.

about 4 in long, containing a raised ring round the circumfere of the base to remove the H_{01}–E_{11} degeneracy. The van silver-plated rod 0·095 in in diameter and 1·275 in long, suspended 1·5 in from the axis of the box by a 0·06 in-diam polystyrene rod. This rod was attached, outside the cavity w to a stiff wire shaft carrying a small mirror and a metal spider which dipped into an annular pool of liquid for dam the motion of the vane. The suspension was a 6 in lengt No. 49 s.w.g. phosphor-bronze wire attached to a torsion h at its upper end.

(5) CALIBRATION

The torsional constant of the suspension was found observing the period of rotational oscillation of a ¼ in steel attached to its lower end.

In order to find the value of $df_0/d\theta$ for the vane and c over the working range of deflection, the vane was ro manually by the torsion head and the resonant frequency c vity for a series of deflections was measured, using a sufficiently nall input to cause negligible force on the vane.

The Q-factor of the cavity was measured by the R.R.D.E. ho-box Q-factor meter,[7] which equates the rate of decay of e oscillation in the cavity with the decay of voltage across a own parallel *RC* circuit. This part of the calibration is erefore not absolute. However, the Q-factor could have been und in terms of the change of cavity impedance with applied equency and hence measured by a piston and standing-wave-dicator experiment involving measurement of length and time ly. It would have been difficult to equal the accuracy of the -meter by this method because of the high Q-factor of the sonator.

(6) OPERATION

When making a measurement, power is fed to the cavity rough a slot coupling whose length is adjusted so that the sonant cavity appears as a matched load to the guide. The ston is then used to tune the cavity to resonance. When the ld inside builds up, the vane moves, thereby lowering the sonant frequency of the cavity. In this condition the vane is a stable state. If it returns towards its zero position, the rce acting on it increases as it brings the box nearer to resonance; motion in the other direction is restrained by the suspension. ırther tuning is required to obtain a greater deflection.

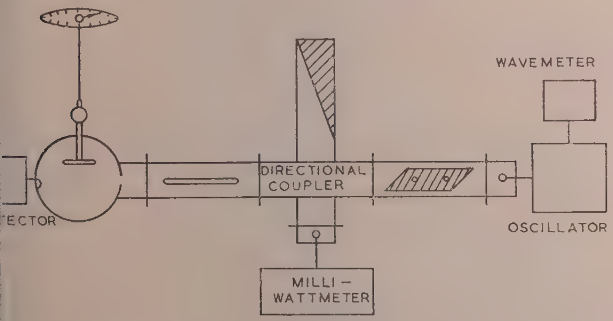
In the final position the torque in the suspension is equal to e maximum torque that can be exerted by the field with the vity fully at resonance. At this point the vane becomes stable; movement of the vane towards its zero position tunes the cavity, thus reducing the deflecting torque, and the ne swings back to zero.

A measurement therefore consists in tuning the cavity slowly obtain the maximum vane deflection and calculating the wer from the values of torque, *Q* and $df_0/d\theta$ at that vane sition.

(7) EXPERIMENTAL RESULT

(7.1) Measurements

A sketch of the arrangement used to check the experimental curacy of the cavity wattmeter is given in Fig. 3.



. 3.—Apparatus used for testing the experimental cavity wattmeter.

The power absorbed in the cavity was measured by means of lirectional coupler of known power division and a thermistor lliwattmeter which had been calibrated against a water orimeter. Simultaneous measurements of the power indi-ed by the milliwattmeter and by the cavity wattmeter were de at various power levels.

In Table 1 the results of eleven such measurements are en.

Table 1

POWER ABSORBED; OBSERVED AND CALCULATED

P_w	D	T	t_0	P_c
mW	cm	dyne-cm $\times 10^{-3}$	μs	mW
24·2	5·4	8·84	3·52	22·04
24·2	5·8	9·50	3·52	23·75
25·6	6·1	10·00	3·52	24·85
28·35	7·4	12·12	3·50	30·4
34·59	8·2	13·4	3·50	33·6
42·4	10·5	17·2	3·48	43·4
48·4	11·4	18·7	3·48	47·1
49·7	11·4	18·7	3·48	47·1
54·6	12·34	20·2	3·46	51·25
63·6	14·8	24·25	3·42	62·3
67·5	15·4	25·22	3·42	64·7

Where P_w = Power absorbed in the cavity as indicated by the milliwattmeter and coupler.
D = Deflection at 67 cm radius.
T = Torque.
t_0 = Unloaded time-constant of the cavity.
P_c = Power absorbed in the cavity, calculated from the formula

$$P_c = \frac{2T}{t_0}\frac{f}{df/d\theta} \times 10^{-1} \quad . \quad . \quad . \quad . \quad (2)$$

The denominator $df/d\theta$ was assumed constant over the range of deflections used and was equal to 63·8 Mc/s per radian.

The sensitivity of the instrument in this experiment was about 0·4 dyne-cm/watt. The values of P_w and P_c are plotted in Fig. 4, where it may be seen that the cavity wattmeter reads true power within a few per cent.

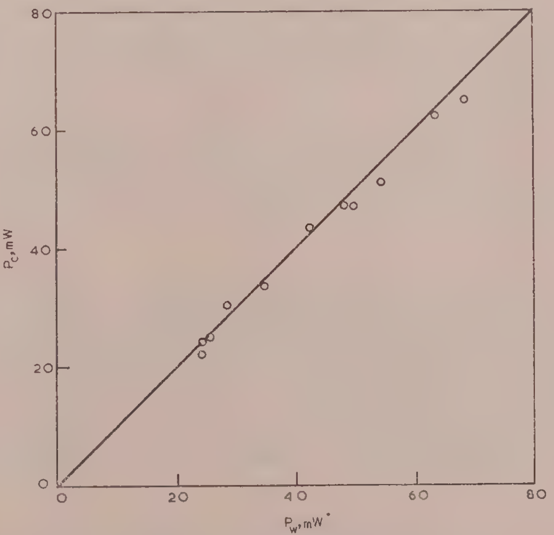


Fig. 4.—Power indication of the cavity wattmeter (P_c) plotted against power indicated by the calibrated milliwattmeter (P_w).

(7.2) Errors

The possible systematic errors amounted to about 7½%. The sum of the errors in the calibration of the milliwattmeter and the directional coupler was about 3½%, while the measurement of the constants of the cavity wattmeter, t_0, $df_0/d\theta$ and the specific torque of the suspension, added about 4%.

The large random errors were chiefly due to the difficulty of tuning the cavity without upsetting the delicate equilibrium of the vane.

(8) FUTURE DEVELOPMENT

Although the power meter has been tested at 3 000 Mc/s only, the principle may be applied at any other frequency, and in practice its main use may be at Q-band or at frequencies of the order of a few hundred megacycles per second where convenient power standards are lacking.

There are a number of difficulties in constructing a practical instrument on the resonant-circuit principle. In the form described in the paper, for instance, the instability in the vane deflection makes the tuning of the cavity, to obtain maximum deflection, very tedious. Furthermore, the bandwidth of the cavity is small because of the high Q-factor, and the device has the undesirable property of presenting a rapidly varying reactive load to the power source.

The cause of the instability of the vane is analysed in Section 11.2, where it is shown that the instability exists only if the vane deflection exceeds certain limits. It is proposed to increase the effective suspension stiffness and to measure the deflecting torque electrically by means of a servo mechanism attached to the vane. The necessary increase in stiffness is of the order of 100 times. With this refinement the cavity could be tuned rapidly to obtain the maximum torque on the vane, and, because of the very small vane movement, changes in Q and $df_0/d\theta$ with vane rotation would be insignificant.

The sensitivity of the instrument is proportional to the product of Q and $df_0/d\theta$, and in a practical wattmeter it would be preferable to work with a low Q-factor and a high value of $df_0/d\theta$ to obtain a larger operating bandwidth. For a loaded Q-factor of 1 000 at 3 000 Mc/s, the force on the vane is within 1% of its maximum value over a 300 kc/s band, which is adequate for most applications.

This reduction in Q-factor will also reduce the rate at which the cavity reactance changes with frequency, but to measure the power output of any source requiring a matched load an attenuating pad between the generator and cavity would be necessary.

(9) CONCLUSION

A sensitive method of power measurement at microwave frequencies has been demonstrated in which calibration does not involve comparison with a d.c. measurement but is derived from simple measurements of

(*a*) The Q-factor of a cavity.

(*b*) The torque on a vane in the cavity.

(*c*) The rate of change of resonant frequency of the cavity with rotation of the vane.

(10) ACKNOWLEDGMENTS

The author gratefully acknowledges contributions to the work described made by Mr. C. M. Burrell and other members of R.R.E., and by Dr. Cullen of University College, London. The paper is published with the permission of the Chief Scientist, Ministry of Supply, and the Controller of H.M. Stationery Office.

(11) REFERENCES

(1) CARRARA, N., and LOMBARDINI, P.: "Radiation Pressure of Centimetre Waves." *Nature*, 1949, **163**, p. 171.

(2) CULLEN, A. L.: "Absolute Power Measurement at Microwa Frequencies," *Proceedings I.E.E.*, Monograph No. 23 February, 1952 (**99**, Part IV, p. 100).

(3) CULLEN, A. L.: "A General Method for the Absolu Measurement of Microwave Power," *ibid.*, Monogra No. 24 M, February, 1952 (**99**, Part IV, p. 112).

(4) CULLEN, A. L., and STEPHENSON, I. M.: "A Torque Operat Wattmeter for 3-cm Microwaves," *ibid.*, Monogra No. 42 M, July, 1952 (**99**, Part IV, p. 294).

(5) MACLEAN, W. R.: "The Resonator Action Theoren *Quarterly Journal of Applied Mathematics*, 1945, **2**, p. 3

(6) SLATER, J. C.: "Microwave Electronics" (Van Nostran 1950).

(7) BURRELL, C. M., and SHAWE, L. W.: "A Q-Factor Co parator for Echo-Boxes in the 10-cm Band," *Jour I.E.E.*, 1946, **93**, Part IIIA, p. 1443.

(12) APPENDICES

(12.1) Derivation of Eqn. (1a), in a Simple Case

Eqn. (1*a*) for the force on a variable reactive element in resonant system can be derived easily for a simple *LCR* circu

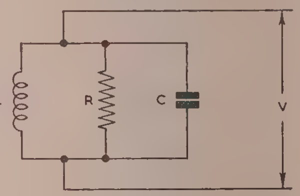


Fig. 5.—Simple *LCR* circuit.

In Fig. 5, let C be an idealized capacitor, so that

$$C = \frac{\epsilon_0 A}{x}$$

where A is the plate area and x is the plate spacing.

Then $$f_0 \propto \frac{1}{\sqrt{C}}$$

$$= K'\sqrt{x} \quad \text{where } K' \text{ is a constant.}$$

Hence $$\frac{df_0}{dx} = \frac{K'}{2\sqrt{x}}$$

$$= \frac{f_0}{2x} \quad . \quad . \quad . \quad . \quad . \quad . \quad .$$

Also $$t_0 = \frac{Q}{\pi f} = \frac{\omega CR}{\pi f}$$

from which $$\frac{t_0}{2} = CR \quad . \quad . \quad . \quad . \quad . \quad . \quad .$$

The power absorbed in the circuit is

$$P = \frac{\overline{V^2}}{R} \quad . \quad . \quad . \quad . \quad . \quad . \quad .$$

Again assuming an idealized capacitor with no fringing f the force on the plate is

$$F = \frac{\epsilon_0}{2} E^2 A$$

where E is the electric field strength.

The mean force on the plates is therefore

$$F = \frac{\epsilon_0}{2}\frac{\overline{V^2}}{x^2}A$$

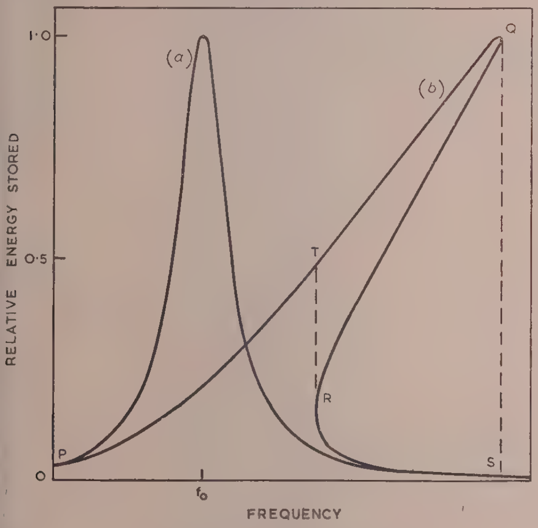
$$= \frac{C}{2}\frac{\overline{V^2}}{x}$$

$$= \frac{\overline{V^2}}{R}CR\frac{1}{f_0}\frac{f_0}{2x} \text{ in the direction of decreasing } x,$$

$$= -P\frac{t_0}{2}\frac{1}{f_0}\frac{df_0}{dx} \text{ from eqns. (3), (4), and (5).}$$

(12.2) Limits of Vane Deflection for Stability

It has been mentioned in Section 6 that the vane deflection sses through a region of instability as the cavity is tuned ough resonance. Since this would be undesirable in a practical wer meter, it is desirable to find the limits of vane deflection thin which no instability occurs.



ig. 6.—Normal and distorted frequency responses of a cavity containing a movable element.

n Fig. 6, curve (*a*) is the normal resonance curve of the cavity, h the vane fixed at its zero position, showing stored energy *W* tted against frequency *f*. If the vane is free to move, however, vill be deflected by an angle proportional to *W*, and since, r a small range, the change of resonant frequency is protional to the angle of deflection, each point on curve (*a*) will displaced along the frequency axis by an amount proportional *W*. The resultant effective resonance curve of the cavity and e is then given by curve (*b*).

)ver the region P–Q on this curve the vane is in stable equiium, since movement towards its zero increases the deflecting ce by bringing the cavity nearer to resonance, and movement the other direction increases the restoring force of the sussion. Over the region R–Q, however, it is unstable because deflecting torque increases with vane rotation at a greater than the suspension torque, and if an attempt is made to trace curve from S towards R, the vane becomes unstable in the vicinity of R and jumps to a point such as T. Similarly, near Q the vane will swing suddenly to point S.

At all stable points on the curve the torque in the suspension, T_s, is equal to the torque exerted by the field, T_f. Instability will occur if the angular rate of change of T_f is greater than the angular rate of change of T_s. For vane stability we require

$$\frac{dT_f}{T_f} < \frac{dT_s}{T_s}$$

i.e.

$$\frac{dW}{W} < \frac{d\theta}{\theta}$$

$$\frac{1}{W}\frac{dW}{df_0}\frac{df_0}{d\theta} < \frac{1}{\theta}$$

or

$$\theta\frac{df_0}{d\theta} < \frac{1}{\frac{1}{W}\frac{dW}{df_0}}$$

The frequency-dependence of stored energy, *W*, is of the form

$$W = \frac{1}{1+x^2} \quad \text{where } x = Q\left(1 + \frac{f^2}{f_0^2}\right)$$

$$\frac{1}{W}\frac{dW}{df_0} = \frac{-2x}{1+x^2}\frac{dx}{df_0}$$

$$= \frac{2x}{1+x^2}\frac{2Q}{f_0}\frac{f^2}{f_0^2}$$

$$\simeq \frac{2x}{1+x^2}\frac{2Q}{f_0}$$

over the region in which we are interested. The expression on the right is easily shown to have a maximum value when $x = 1$.

Then

$$\frac{1}{W}\frac{dW}{df_0} = \frac{2Q}{f_0}$$

The requirement for a stable system is

$$\theta\frac{df_0}{d\theta} < \frac{f_0}{2Q}$$

i.e. the maximum frequency deviation from f_0 due to vane movement must be less than half the bandwidth of the resonant circuit.

When the cavity is fed from a finite source-impedance, the relative bandwidth is the loaded bandwidth of the cavity, and in the particular case of a matched cavity it is $2f_0/Q$.

In the H_{01} cavity, Q (unloaded) $\simeq$ 30 000, and therefore the loaded bandwidth was about 200 kc/s. The maximum deflection of the vane for a power input of 50 mW was 0·085 rad.

Therefore

$$\theta\frac{df_0}{d\theta} = 0{\cdot}085 \times 63{\cdot}8\,\text{Mc/s}$$

$$= 5\,420\,\text{kc/s}.$$

An increase of over 50 times in the vane suspension stiffness would have been required to make the vane stable.

In the case of the variable capacitor in the lumped circuit of Section 11.1, the stability condition may be shown to be $\Delta x < (1/Q)x$, where Δx is the maximum plate movement, and x is the plate spacing.

621.372.4/5:621.3.011

Monograph No. 1
July

THE CORRELATION BETWEEN DECAY TIME AND AMPLITUDE RESPONSE

By S. DEMCZYNSKI, Dipl.Ing., Graduate.

(*The paper was first received* 1*st November*, 1954, *and in revised form* 11*th March*, 1955. *It was published as an* INSTITUTION MONOGRA *in July*, 1955.)

SUMMARY

The object of the present work is to investigate the correlation existing between the decay time and the delay time of the indicial response on the one hand, and the bandwidth and peak values of the steady-state amplitude response and the slope of the phase response on the other.

The analysis is applied to five types of electrical networks and to some general classes of circuits, distinguished by the location of their poles in the *p*-plane. From these investigations the following results are obtained: (*a*) By interpolation of all numerical results obtained for the five types of networks, the statistical formulae embracing all of them are found. (*b*) Formulae giving the functional relationship between the decay time and the ratio f_3/f_6 are derived for each class of circuit considered. (*c*) The formula for the delay time, $t_l = [d\phi/d\omega]_{\omega=0}$, is checked for all five networks considered and is found to be valid for multi-stage, but not for single-stage, circuits.

LIST OF PRINCIPAL SYMBOLS

f_0 = Resonant frequency of the circuit.
f_3 = Bandwidth of amplitude response at 3 dB level below that at centre frequency.
f_6 = Bandwidth of amplitude response at 6 dB level below that at centre frequency.
f_3' = Bandwidth of amplitude response of one stage at 3 dB level below centre frequency.
$\mathbf{H}(f)$ = Normalized frequency response.
$H(f)$ = Normalized amplitude response.
H_1 = Maximum value of normalized amplitude response.
n = Number of stages.
$Q = 2\pi f_0 L/R$
$t_d(1\%)$ = Decay time.
$t_d'(1\%)$ = Total decay time.
t_l = Delay time.
t_l' = Slope of phase response at centre frequency.
$v(t)$ = Indicial response of filter.
α = Real co-ordinate of pole in *p*-plane.
β = Imaginary co-ordinate of pole in *p*-plane.
$\gamma(1\%) = f_6 t_d(1\%)$.
$\gamma'(1\%) = f_6 t_d'(1\%)$.
$\sigma = \dfrac{f_3}{f_6} H_1$
$\phi(f)$ = Phase response.

(1) INTRODUCTION

The aim of the present work is to investigate the correlation existing between decay time and delay time of the indicial response of certain types of four-terminal networks on the one hand, and, on the other, certain easily measurable parameters of their normalized amplitude response, namely the ratio f_3/f_6 and the peak value H_1 (see Fig. 1).

The amplitude response is the modulus of the frequency response, which is defined as the complex ratio of the output and input voltages. It represents the performance of the sy under steady-state conditions, when its behaviour can be desc by functions which are constant or periodic in time at con frequency, and may be normalized by taking its value a centre frequency as unity.

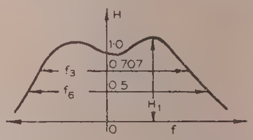


Fig. 1.—Steady-state amplitude characteristic.

The transient period may be considered as the time in w the system passes from one steady state of energy conditio the other. In order to obtain more uniform results and a b basis for comparison of various networks, the transients sidered in the present work are all responses to the unit vo step. Such transients are called "indicial responses."

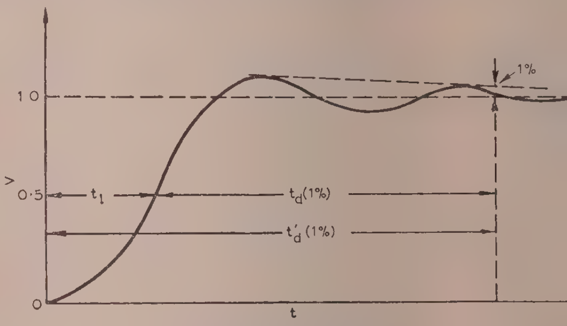


Fig. 2.—Indicial response.

The decay time (see Fig. 2) is defined as the time taken the mid-point amplitude of the indicial response to the in when the envelope of the indicial response deviates fron steady state by a prescribed proportion of the amplitude $t_d(1\%)$ or $t_d(0\cdot1\%)$. The delay time is defined as the time by the indicial response curve to reach the half-amplitude p

The circuits considered are all minimum-phase passive works consisting of lumped parameters and comprising low-pass and band-pass filters. The band-pass filter havi amplitude response very nearly identical with a particular of a low-pass filter, but shifted by f_0 on the frequency a called a band-pass analogue of that low-pass filter. The res of the low-pass filter to the unit voltage step is identical wi envelope of the response of its band-pass analogue to th step of carrier voltage at frequency f_0.

It is known from general circuit theory that any p minimum-phase network can be unambiguously described by its amplitude response or its indicial response. It is easy to measure or calculate the steady-state response of a filter, and the influence of particular parameters on the sh

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Demczynski is with Marconi's Wireless Telegraph Co., Ltd.

ıe amplitude responses can usually be readily discovered. The arious methods of synthesis of electrical filters for specified mplitude responses employ the wealth of analytical and mpirical research begun last century.

The techniques for designing a system for a required transient erformance are much less developed. In order to find the ıdicial response of a network, it is necessary to solve the fundaıental differential equations describing its behaviour, or obtain ıe inverse Fourier or Laplace transform of the steady-state equency characteristic. Either method is very laborious, and ıe particular parameters on the shape of the indicial response is sually extremely difficult to assess.

The transient performance is probably of most importance in ervo mechanisms, and with the spectacular development of this ranch of technology in the last decade, various formulae, mostly mpirical, have been presented giving the relations between the arameters of frequency response on the one hand and the overıoot and rise time of the indicial response on the other.

Similar formulae referring to the other parameters of the ıdicial response are presented here. Some are deduced in ection 2.3 by comparing directly the expressions representing ıe amplitude response and indicial response of the given type f network. Others are found in Section 2.2 by interpolation of ıe results obtained for many types of circuit. The former resent the existing relations more exactly, while the latter are ıore approximate but at the same time simpler and more suitable or the designer to use.

(2) THE DECAY TIME

(2.1) The General Approach

Apart from obvious importance in some servo mechanisms, ıe decay time may also be of interest in certain amplitudeıodulated multi-channel pulse communication systems.

Increasing the slope of the amplitude response results in an ıcrease of the overshoot, which may be considered as the initial mplitude of a quasi-sinusoidal exponentially-damped oscillation ı the "top" of the indicial response. It also reduces the attenuaon of this oscillation. Both effects increase the decay time. Iowever, when there is no oscillation as in the case of n RC stages, ıe decay time, as defined here, is reduced with increasing slope. 'he apparent connection between the decay time and the slope f the amplitude response suggests the possibility of finding a eneral functional relationship between these two quantities.

(2.2) The Statistical Approach

To find the relationship just mentioned five different types of etworks are selected and the decay times for various values of Q and n are calculated. At the same time the corresponding alues of σ are found. The results are presented in Figs. 3 and 4 ı the form $t_d(1\%)f_6 = \gamma = F(\sigma)$. By interpolating all the curves y a single straight line, on the principle of least total relative quare error, the following statistical formulae, valid for all ırcuits considered, are found:

$$\gamma(1\%) = 6\cdot73\sigma - 2\cdot76 \quad . \quad . \quad . \quad . \quad (1)$$

$$\gamma(0\cdot1\%) = 9\cdot25\sigma - 3\cdot26 \quad . \quad . \quad . \quad . \quad (2)$$

Because of the variety of circuits taken into consideration, the pproximate formulae (1) and (2) can be expected to be valid for ıany other passive networks and may well serve as a first rough ıdication for the designer. It can be seen that, in order to ecrease decay time, the slope of the amplitude response must be ecreased. For the same slope, the circuit having an amplitude esponse with peaks will have a longer decay time than one with flat response. The main formulae necessary to obtain

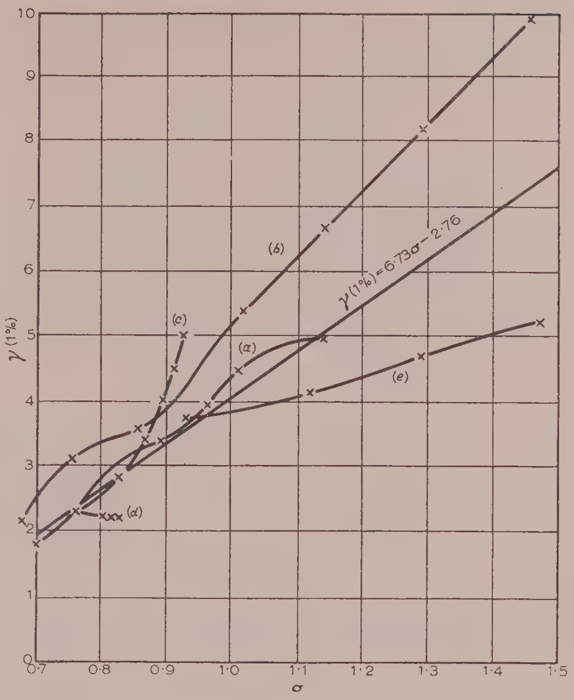


Fig. 3.—Graph of $\gamma(1\%)$ against σ.

(a) Series peaking coil.
(b) Shunt peaking coil.
(c) Staggered circuits.
(d) Critically coupled LRC circuits.
(e) Over-coupled circuits.

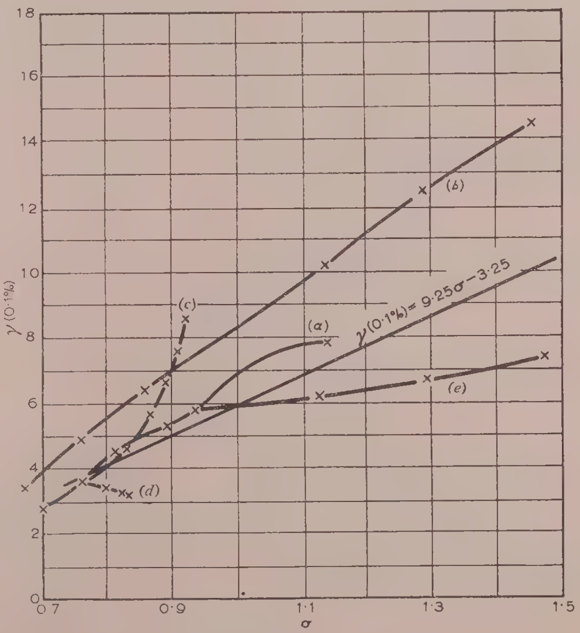


Fig. 4.—Graph of $\gamma(0\cdot1\%)$ against σ.

(a) Series peaking coil.
(b) Shunt peaking coil.
(c) Staggered circuits.
(d) Critically coupled LRC circuits.
(e) Over-coupled circuits.

eqns. (1) and (2) are given in Section 6. Many may be of interest for their own sake, since it is believed that they have not been presented elsewhere.

(2.3) Analytical Approach

(2.3.1) General Discussion.

Another approach to the problem is based on the fact that any passive minimum-phase network is completely defined by the location of its poles and zeros in the p-plane, except for a possible constant multiplier. Hence certain classes of circuits can be distinguished, e.g. filters having frequency response with one, two or more single or multiple poles. An attempt is made here to derive the general formulae, valid for all circuits belonging to a given class, which give the functional dependence in the form

$$t'_d(1\%)f_6 = \gamma'(1\%) = F(f_3/f_6) \quad . \quad . \quad . \quad . \quad (3)$$

where $t'_d(1\%)$ is the total decay time, defined as the time taken from the moment of application of input unit voltage step to the moment when the envelope of the transient response deviates from the steady state by a prescribed proportion of the amplitude, e.g. $t'_d(1\%)$. The reason for introducing this total decay time into the present Section, instead of using the definition of Section 1, is to make analytical formulae like eqn. (3) more simple. On the other hand, it is felt that the decay time as defined in Section 1 may be of more immediate interest to the designer. It should be observed that, although a resonant circuit has two conjugate poles in the p-plane, it is well known that, if the value of the Q-factor is high, only the pole having a positive imaginary co-ordinate contributes effectively to the amplitude response at real frequencies. If a carrier be taken at the resonant frequency of this circuit and modulated by a unit step, the frequency spectrum obtained is concentrated about the frequency corresponding to this pole.

Hence when calculating either the indicial response or amplitude response of the resonant circuit, the pole with negative imaginary co-ordinate may be neglected. Therefore, when referring to poles of single-tuned coupled or staggered circuits in the present Section, only those poles with positive imaginary co-ordinates in the p-plane are considered.

(2.3.2) Circuits with One Single Pole.

The normalized amplitude response is of the form

$$H(f) = \frac{\alpha}{\sqrt{(\alpha^2 + \omega^2)}} \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and the indicial response is of the form

$$v(t) = 1 - \varepsilon^{-\alpha t} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Hence $$\gamma'(1\%) = \frac{\sqrt{(3)}\log_\varepsilon 100}{\pi} = 2\cdot 54 \quad . \quad . \quad . \quad (6)$$

$$\gamma'(0\cdot 1\%) = 3\cdot 82 \quad . \quad . \quad . \quad . \quad . \quad (7)$$

(2.3.3) Circuits with Two Single Poles.

To the class of circuit with two single poles belong, among others, the single-stage series peaking coil and a pair of coupled resonant circuits.

The final formulae are as follows:

$$H(f) = \frac{(\alpha^2 + \beta^2)}{\{[\alpha^2 + (\omega - \beta)^2][\alpha^2 + (\omega + \beta)^2]\}^{\frac{1}{2}}} \quad . \quad (8)$$

$$v(t) = 1 - \mathscr{R}\left[\left(\frac{1}{\beta}\right)\sqrt{(\alpha^2 + \beta^2)}\varepsilon^{-\alpha t}\varepsilon^{j(\beta t + \theta)}\right] \quad . \quad . \quad (9)$$

$$\gamma'(1\%) = \frac{2}{\pi}\left(\frac{3x^2 - 1}{1 - 3x^4 + y}\right)^{1/2} \log_\varepsilon 100\left(\frac{2y}{3x^4 - 1 + y}\right)^{1/2} \quad . \quad . \quad . \quad . \quad (1$$

$$\gamma'(0\cdot 1\%) = \frac{2}{\pi}\left(\frac{3x^2 - 1}{1 - 3x^4 + y}\right)^{1/2} \log_\varepsilon 1000\left(\frac{2y}{3x^4 - 1 + y}\right)^{1/2} \quad . \quad . \quad . \quad . \quad (1$$

where $$x = f_3/f_6 \quad . \quad . \quad . \quad . \quad . \quad . \quad (1$$

$$y = 2x\sqrt{[(1 - x^2)(3x^2 - 1)]} \quad . \quad . \quad . \quad (1$$

Eqns. (10) and (11), when applied to coupled circuits, a valid for any degree of coupling, whether the amplitude respor has peaks or not. Hence, so far as decay time is concerne the passage from undercoupled to overcoupled circuits is co tinuous, and the appearance of the peaks does not introdu anything essentially different.

(2.3.4) Circuits having Three Single Poles.

The investigations carried out on circuits having three sin poles lead to the conclusion that, in order to obtain formul analogous to eqns. (10) and (11), a set of three cubic equatio with three unknowns must be solved. The solution leads extremely complicated formulae, unsuitable for practical u

(2.3.5) Staggered Circuits.

Staggered circuits have frequency responses with n poles a no zeros. However, the location of the poles is not arbitrar they are all the 2nth roots of $(-1)^{n+1}$ and lie to the left of t imaginary p-axis on the unit semi-circle. Therefore n determin unequivocally the frequency and indicial response, and hen also the total decay time.

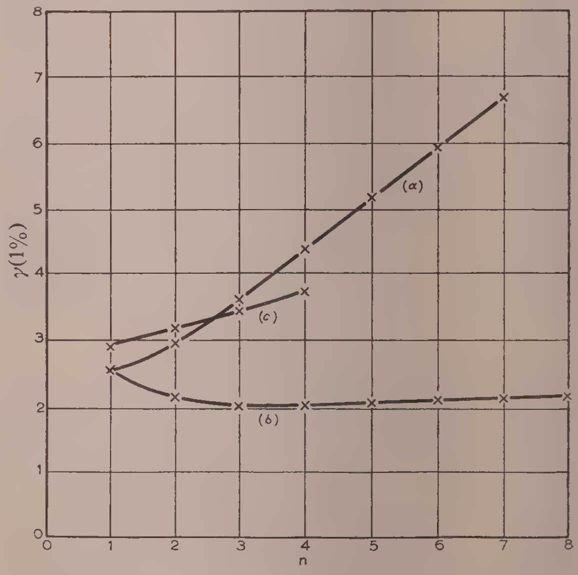


Fig. 5.—Graph of $\gamma'(1\%)$ against number of stages n.

(a) Staggered circuits.
(b) Circuits with one multiple pole.
(c) Critically coupled circuits.

Approximate formulae are as follows:

$$\gamma'(1\%) = \frac{3^{1/2n}}{\pi \sin\frac{\pi}{2n}} \log_\varepsilon 100 \frac{1}{2^{n-2}\prod_{k=1}^{k=n-1}\sin\frac{\pi}{2n}k} \quad . \quad (1$$

$$n \geqslant 2$$

$$\gamma'(0\cdot1\%) = \frac{3^{1/2n}}{\pi\sin\frac{\pi}{2n}}\log_\varepsilon 1\,000\frac{1}{2^{n-2}\prod\limits_{k=1}^{k=n-1}\sin\frac{\pi}{2n}k} \quad . \quad (15)$$

To facilitate their use, the above formulae are presented in raphical form as curves (*a*) in Figs. 5 and 6.

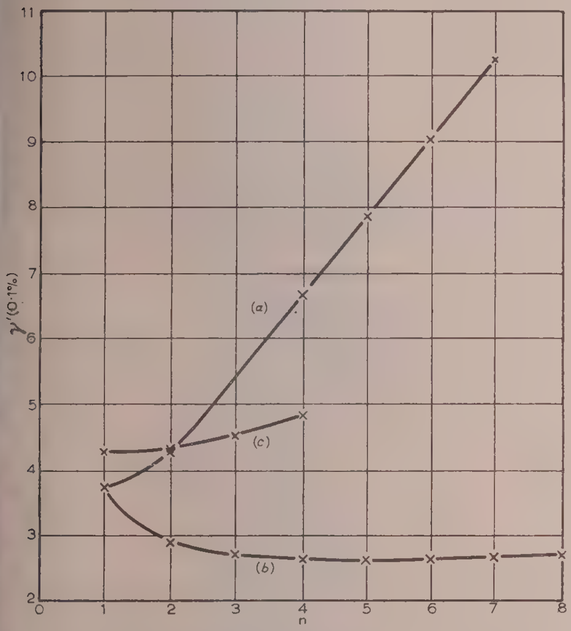


Fig. 6.—Graph of $\gamma'(0\cdot1\%)$ against number of stages *n*.
(*a*) Staggered circuits.
(*b*) Circuits with one multiple pole.
(*c*) Critically coupled circuits.

2.3.6) Circuit with One Multiple Pole.

The circuit with one multiple pole is represented by *n RC* ircuits or by its band-pass analogue, *n LRC* identical tuned ircuits. For this circuit $\gamma'(1\%)$ and $\gamma'(0\cdot1\%)$ are functions of he number of stages only, and these relationships may be put nto the form

$$\varepsilon^{\gamma'(1\%)f(n)} = 100\sum_{r=0}^{r=n-1}\frac{[\gamma(1\%)f(n)]^{n-r-1}}{(n-r-1)!} \quad . \quad . \quad (16)$$

$$\varepsilon^{\gamma'(0\cdot1\%)f(n)} = 1\,000\sum_{r=0}^{r=n-1}\frac{[\gamma(0\cdot1\%)f(n)]^{n-r-1}}{(n-r-1)!} \quad . \quad (17)$$

here $$f(n) = \frac{\pi}{(2^{2/n}-1)^{1/2}} \quad . \quad . \quad . \quad . \quad . \quad (18)$$

There is no dominant term in either of the expressions, and onsequently it seems impossible to approximate the value of γ' y any reasonably simple formula. Hence the values of γ' were ound by approximate numerical solution of the exponential qns. (16) and (17) for the consecutive integral values of *n*.

The results of these calculations are shown by curves (*b*) in igs. 5 and 6.

.3.7) Circuits with Two Multiple Poles.

To the class of circuits with two multiple poles belong, among thers, networks consisting of *n* stages of coupled circuits with ny degree of coupling. If all terms of the indicial response are aken into account, it appears readily that the derivation of any ractically useful formula can hardly be expected.

However, if only over-coupled circuits are considered, with critically coupled circuits as the limit, it can be shown that the indicial response is approximated fairly accurately by its dominant term, and the following formulae can be derived:

$$\gamma'(1\%) = \frac{2}{\pi}\left(\frac{ax^2-b}{b-ax^4+y}\right)^{1/2}\log_\varepsilon\left\{\frac{100\pi^{n-1}}{(n-1)!2^{n-1}}\gamma'(1\%)^{n-1}\right.$$
$$\left.\left[\frac{x^2(1-x^2}{ax^2-b}\right]^{(n-1)/4}\left[\frac{2y}{ax^4-b+y}\right]^{n/2}\right\} \quad . \quad (19)$$

$$\gamma'(0\cdot1\%) = \frac{2}{\pi}\left(\frac{ax^2-b}{b-ax^4+y}\right)^{1/2}\log_\varepsilon\left\{\frac{1\,000\pi^{n-1}}{(n-1)!2^{n-1}}\gamma'(0\cdot1\%)^{n-1}\right.$$
$$\left.\left[\frac{x^2(1-x^2}{ax^2-b}\right]^{(n-1)/4}\left[\frac{2y}{ax^4-b+y}\right]^{n/2}\right\} \quad . \quad (20)$$

where

$$x = f_3/f_6 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

$$y = 2x\sqrt{[(1-x^2)(ax^2-b)]} \quad . \quad . \quad . \quad . \quad (22)$$

$$a = 2^{2/n}-1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

$$b = 2^{1/n}-1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

Eqns. (19) and (20) are presented in graphical form in Figs. 7 and 8 for various values of the parameter *n*. For $n = 1$, these formulae reduce to eqns. (10) and (11) respectively.

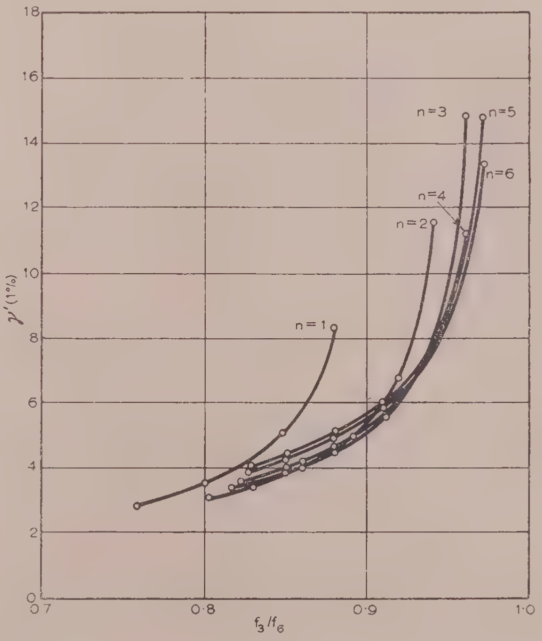


Fig. 7.—Graph of $\gamma'(1\%)$ against f_3/f_6 for circuits with two multiple poles.

For critically coupled çircuits, eqris. (19) and (20) reduce respectively to

$$\gamma'(1\%) = \frac{(4a)^{1/4}}{\pi}\log_\varepsilon\left[\frac{100\pi^{n-1}}{(n-1)!2^{n-1}}\gamma'(1\%)^{n-1}\frac{2^{n/2}}{a^{(n-1)/4)}}\right] \quad (25)$$

$$\gamma'(0\cdot1\%) = \frac{(4a)^{1/4}}{\pi}\log_\varepsilon\left[\frac{1\,000\pi^{n-1}}{(n-1)!2^{n-1}}\gamma'(0\cdot1\%)^{n-1}\frac{2^{n/2}}{a^{(n-1)/4}}\right] \quad (26)$$

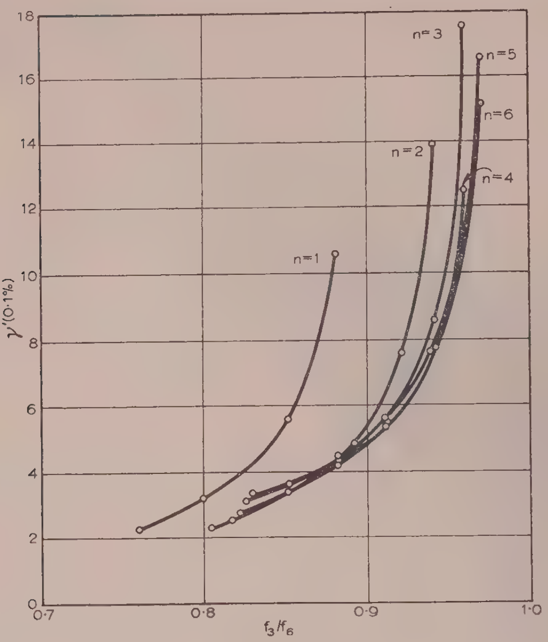


Fig. 8.—Graph of $\gamma'(0\cdot 1\%)$ against f_3/f_6 for circuits with two multiple poles.

which give in implicit form the total decay time as a function of the number of stages only. Eqns. (25) and (26) are represented by curves (*c*) in Figs. 5 and 6.

The error introduced by eqns. (14), (15), (16), (17), (19), (20), (25) and (26) is a few per cent only, and hence they can be safely used for practical purposes.

In eqns. (19) and (20) the parameter *n* and the slope f_3/f_6 appear. The parameter *n* cannot be eliminated by means of f_3/f_6, because the same slope of amplitude response can be obtained for different numbers of stages, if for each value of *n* suitable values of α and β are found.

This is obvious from the easily derivable formula

$$f_3/f_6 = \left\{\frac{\beta^2 - \alpha^2 + \sqrt{[(\beta^2 - \alpha^2) + (\alpha^2 + \beta^2)b]}}{\beta^2 - \alpha^2 + \sqrt{[(\beta^2 - \alpha^2) + (\alpha^2 + \beta^2)a]}}\right\}^{1/2} \qquad (27)$$

where *a* and *b* are defined by eqns. (23) and (24) respectively. The above formula is valid for *n* stages of coupled circuits.

(3) DELAY TIME

If to a filter having a frequency characteristic

$$H(f) = H(f)\varepsilon^{j\phi(f)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

an input voltage pulse having spectrum *G*(*f*) is applied, the output voltage is given by

$$v(t) = \int_{-\alpha}^{\alpha} G(f)H(f)\varepsilon^{j[2\pi ft+\phi(f)]}df \quad . \quad . \quad . \quad (29)$$

Hence every frequency component of the output voltage is delayed by

$$\Delta t = \frac{\phi(f)}{2\pi f} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

with respect to the corresponding component of input pulse. Thus, in contrast to the decay time, which is more obviously connected with the amplitude response, the delay time seems t be more readily correlated with the phase response.

From the character of the spectrum of the unit voltage ste it is evident that the bulk of the energy is transmitted by th components near to the zero frequency. This effect is increase by the discrimination introduced by the filter itself. In th frequency range occupied by these components, the phas response of the filter is usually almost linear; and hence the delay time, which may be regarded approximately as the dela time of the whole pulse, is given by

$$t'_l = \left(\frac{d\phi}{d\omega}\right)_{\omega=0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3$$

Eqn. (31) has been checked for all circuits previously considered by comparing in Table 1 the results given by eqn. (3 with the exact values of t_l obtained by the means indicated Section 6.

Table 1

Comparison of t_l and t'_l

Q	Series peaking coil		Shunt peaking coil	
	$\pi f_3 t_l$	$\pi f_3 t'_l$	$\pi f_3 t_l$	$\pi f_3 t'_l$
0·6	1·54	5·4	0·98	4·8
0·707	1·43	2·89	0·78	2·19
0·8	1·38	2·05	0·67	1·25
1	1·29	1·33	0·55	0·39
1·1	1·26	1·17	0·5	0·07
1·2	1·25	1	0·43	0·05

n	Critically-staggered circuits		Critically-coupled circuits		Over-coupled circuits	
	$\pi f_3 t_l$	$\pi f_2 t'_l$	$\pi f'_3 t_l$	$\pi f'_3 t'_l$	$\pi f'_3 t_l$	$\pi f'_3 t'_l$
1	0·7	1	1·47	1·41	1·31	1·044
2	1·47	1·41	3·02	2·82	2·63	2·088
3	2·13	2	4·4	4·23	3·92	3·132
4	2·82	2·42	6·0	5·64	5·2	4·176
5	3·51	3·23				
6	4·14	3·88				
7	4·83	4·44				

It can be seen from Table 1 that eqn. (31) gives an adequa approximation for multiple-stage circuits, but it seems to b quite wrong for the single-stage circuits here considered. Th reason for this may be that in the single-stage circuit the ha monics lying outside the linear range of phase response a rather poorly attenuated, and the previously explained idea delay time has no meaning, so that eqn. (31) has no justificatic in this case.

The degree of approximation given by eqn. (31) in the case over-coupled circuits is lower than in the case of critical coupled circuits; this may be because, in the first case, th frequency-spectrum components corresponding to the peaks the amplitude response are magnified by the filter. Because the components are delayed more than those near the middle fr quency, the whole output pulse is delayed more than wou appear from eqn. (31).

(4) ACKNOWLEDGMENTS

The paper is based on the author's Diploma Thesis at th Polish University College in 1951. The work was carried o under the supervision of Prof. Z. Jelonek, to whom ackno ledgment is due for many helpful suggestions, and in particul for suggesting the statistical approach to the problem. The author also wishes to acknowledge the assistance which he has received from Marconi's Wireless Telegraph Co., Ltd., in the final preparation of the paper.

(5) REFERENCES

(1) GUILLEMIN, E. A.: "Communication Networks," Vol. II (Wiley, 1945).

(2) BODE, H.: "Network Analysis and Feedback Amplifier Design" (Van Nostrand, 1945).

(3) WALIMAN, H.: "Vacuum Tube Amplifiers," M.I.T. Radiation Laboratory Series Vol. 17 (McGraw-Hill, 1947).

(4) CARSLAW, H. S., and JAEGER, J. C.: "Operational Methods in Applied Mathematics" (Oxford University Press, 1941).

(5) GARDNER, M., and BARNES, J.: "Transients in Linear Systems" (Wiley, 1942).

(6) LEVY, M.: "The Impulse Response of Electrical Networks," *Journal I.E.E.*, 1943, **90**, Part III, p. 153.

(7) JELONEK, Z.: "Transient Response," Lectures (unpublished).

(8) KALIMAN, H. E., and SPENCER, R. E.: "Transient Response," *Proceedings of the Institute of Radio Engineers*, 1945, **31**, p. 169.

(9) JAWORSKI, Z.: Diploma Thesis, 1950, Polish University College.

(6) APPENDICES

(6.1) Series Peaking Coil

The series peaking coil is a low-pass filter which has the pair of coupled circuits for its band-pass analogue. The study of the behaviour of either is equivalent to the study of the other. Hence the influence of the Q-factor of the circuit on the decay time is studied for the series peaking coil, whereas the effect of multiple states is dealt with for coupled circuits.

The general expression[8] for the indicial response is as follows:

$$v(t) = 1 - \sqrt{\left(\frac{4Q^2}{4Q^2-1}\right)}\varepsilon^{-\omega_0^t/2Q}$$

$$\cos\left[\sqrt{\left(\frac{4Q^2-1}{4Q^2}\right)}\omega_0 t - \sin\left(\frac{1}{2Q}\right)\right] \quad . \quad (32)$$

$$\gamma(1\%) - \frac{1}{\pi}\sqrt{\left\{\left(1-\frac{1}{2Q}\right) + \left[\left(\frac{1}{2Q}-1\right)^2+3\right]^{1/2}\right\}}$$

$$\left\{\frac{2Q}{0\cdot4343}\left[2+\frac{1}{2}\log_{10}\frac{4Q^2}{4Q^2-1}\right] - t_l\omega_0\right\} \quad . \quad (33)$$

$$\gamma(0\cdot1\%) = \frac{1}{\pi}\sqrt{\left\{\left(1-\frac{1}{2Q}\right) + \left[\left(\frac{1}{2Q}-1\right)^2+3\right]^{1/2}\right\}}$$

$$\left\{\frac{2Q}{0\cdot4343}\left[3+\frac{1}{2}\log_{10}\frac{4Q^2}{4Q^2-1}\right] - t_l\omega_0\right\} \quad . \quad (34)$$

The quantity $\omega_0 t_l$ is found by trial and error for each value of Q from the equation

$$\omega_0 t_l = \frac{2Q}{0\cdot4343}\left\{\log_{10} 2 + \frac{1}{2}\log_{10}\frac{4Q^2}{4Q^2-1}\right.$$

$$\left. + \log_{10}\cos\left[\frac{\sqrt{(4Q^2-1)}}{4Q^2}\omega_0 t_l - \text{arc sin}\left(\frac{1}{2Q}\right)\right]\right\} \quad . \quad . \quad . \quad . \quad (35)$$

and

$$\sigma = +\left[\frac{\sqrt{(1-4Q+8Q^2)}-1+2Q}{\sqrt{(1-4Q+16Q^2)}-1+2Q}\right]^{1/2}\left[\frac{1}{Q^2}-\frac{1}{4Q^2}\right]^{1/2} \quad (36)$$

(6.2) Shunt Peaking Coil

With the shunt peaking coil the voltage on the capacitor starts to rise immediately, because it is not delayed by a coil as in the previous circuits. The rise time decreases with increasing Q-factor. The formulae of interest are as follows:

$$v(t) = 1 - \sqrt{\left(\frac{4Q^4}{4Q^2-1}\right)}\varepsilon^{-\omega_0 t/2Q}$$

$$\cos\left[\sqrt{\left(\frac{4Q^2-1}{4Q^2}\right)}\omega_0 t - \text{arc sin}\left(\frac{2Q^2-1}{2Q^2}\right)\right] \quad (37)$$

$$\gamma(1\%) = \frac{1}{\pi}\left\{1+2Q^2-\frac{1}{2Q^2}\right.$$

$$\left. + \sqrt{\left[\left(1+2Q^2-\frac{1}{2Q^2}\right)^2+3\right]}\right\}^{1/2}$$

$$\left[\frac{2Q}{0\cdot4343}\left(2+\frac{1}{2}\log_{10}\frac{4Q^4}{4Q^2-1}\right) - \omega_0 t_l\right] \quad (38)$$

$$\gamma(0\cdot1\%) = \frac{1}{\pi}\left\{1+2Q^2-\frac{1}{2Q^2}\right.$$

$$\left. + \sqrt{\left[\left(1+2Q^2-\frac{1}{2Q^2}\right)^2+3\right]}\right\}^{1/2}$$

$$\left[\frac{2Q}{0\cdot4343}\left(3+\frac{1}{2}\log_{10}\frac{4Q^4}{4Q^2-1}\right) - \omega_0 t_l\right] \quad (39)$$

$$f_3/f_6 = \left[\frac{a+\sqrt{(a^2+1)}}{b+\sqrt{(a^2+3)}}\right]^{1/2} \quad . \quad . \quad . \quad . \quad (40)$$

$$a = 1 + Q^2 - \frac{1}{2Q^2} \quad . \quad . \quad . \quad . \quad . \quad (41)$$

$$b = 1 + 2Q^2 - \frac{1}{2Q^2} \quad . \quad . \quad . \quad . \quad (42)$$

(The values for $\omega_0 t_l$ are taken from Reference 8.) The values of H_1 are taken from Reference 9.

(6.3) Staggered Circuits

It is possible to stagger the n single tuned circuits in such a way that the resultant normalized amplitude response is maximally flat.

For n staggered circuits it follows that

$$\sigma = \frac{1}{3^{1/2n}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

$n = 1$

$$v(t) = 1 - \varepsilon^{-\pi f_3 t} \quad . \quad . \quad . \quad . \quad . \quad (44)$$

$n = 2$

$$v(t) = 1 - 1\cdot41\varepsilon^{-0\cdot707 f_3 \pi t}\sin\left(0\cdot707\pi f_3 t - \frac{\pi}{4}\right) \quad . \quad (45)$$

$n = 3$

$$v(t) = 1 - \varepsilon^{-\pi f_3 t} - 1\cdot154\varepsilon^{-0\cdot5\pi f_3 t}\cos\left(0\cdot866\pi f_3 t + \theta_1\right) \quad (46)$$

$n = 4$

$$v(t) = 1 - 1\cdot001\varepsilon^{-0\cdot382\pi f_3 t}\cos\left(0\cdot924\pi f_3 t + \theta_2\right)$$

$$- 2\cdot412\varepsilon^{-0\cdot924\pi f_3 t} \times \cos\left(0\cdot382\pi f_3 t + \theta_3\right) \quad . \quad (47)$$

$n = 5$

$$v(t) = 1 - 1\cdot37\varepsilon^{-\pi f_3 t} - 0\cdot896\varepsilon^{-0\cdot309\pi f_3 t}\cos(0\cdot951\pi f_3 t + \theta_4) - 2\cdot768\varepsilon^{-0\cdot809\pi f_3 t}\cos(0\cdot587\pi f_3 t + \theta_5) \quad . \quad . \quad . \quad . \quad (48)$$

$n = 6$

$$v(t) = 1 - 0\cdot816\varepsilon^{-0\cdot258\pi f_3 t}\cos(0\cdot966\pi f_3 t + \theta_6) - 3\cdot054\varepsilon^{-0\cdot707\pi f_3 t}\cos(0\cdot707\pi f_3 t + \theta_7) - 5\cdot279\varepsilon^{-0\cdot965\pi f_3 t}\cos(0\cdot258\pi f_3 t - \theta_8) \quad . \quad (49)$$

$n = 7$

$$v(t) = 1 - 0\cdot766\varepsilon^{-0\cdot2225\pi f_3 t}\cos(0\cdot975\pi f_3 t + \theta_9) - 3\cdot312\varepsilon^{-0\cdot623\pi f_3 t}\cos(0\cdot782\pi f_3 t + \theta_{10}) - 0\cdot678\varepsilon^{-0\cdot9\pi f_3 t}\cos(0\cdot436\pi f_3 t + \theta_{11}) + 4\cdot312\varepsilon^{-\pi f_3 t} \quad . \quad . \quad . \quad . \quad (50)$$

In these equations, the θ_j are phase angles which arise in the calculations. Their values have not actually been calculated, since they do not affect the final results. In every case for the values of t in question, the dominant term is of two orders greater than the next one. Hence only the dominant terms are taken for the calculation of the decay time, e.g. $n = 7$.

$$\gamma(1\%) = \left(\frac{\log 75\cdot6}{0\cdot4343 \times 0\cdot2225} - \pi f_3 t_l\right)\frac{3^{1/2n}}{\pi} \qquad (51)$$

The values of $\pi f_3 t_l$ are taken from Reference (3).

(6.4) Critically-Coupled Tuned Circuits

For n pairs of critically coupled resonant circuits

$$\sigma = \sqrt[4]{\left(\frac{2^{1/n} - 1}{2^{2/n} - 1}\right)} \quad . \quad . \quad . \quad . \quad . \quad (52)$$

$n = 1$

$$v(t) = 1 - 1\cdot41\varepsilon^{-0\cdot707\pi f_3 t}\cos 0\cdot0707\pi f_3 t \quad . \quad (53)$$

$n = 2$

$$v(t) = 1 + \varepsilon^{-0\cdot707\pi f_3' t}\left[\pi f_3' t \cos\left(0\cdot707\pi f_3' t + \frac{\pi}{4}\right) - \sqrt{5}\cos(0\cdot707\pi f_3' t - \text{arc tan } 2)\right] \quad . \quad (54)$$

$n = 3$

$$v(t) = 1 + \frac{1}{2}\varepsilon^{-0.707\pi f_3' t}\left[\frac{(\pi f_3' t)^2}{\sqrt{2}}\cos\left(0\cdot707\pi f_3' t - \frac{\pi}{4}\right) + \pi f_3' t \times 4\cdot123\cos(0\cdot707\pi f_3' t + 31^\circ) - 7\cdot28\cos(0\cdot707 f_3' t - 74^\circ)\right] \quad . \quad (55)$$

$n = 4$

$$v(t) = 1 + \frac{1}{6}\varepsilon^{-0\cdot707\pi f_3' t}\left[-\frac{(\pi f_3' t)^3}{2}\cos\left(0\cdot707\pi f_3' t + \frac{\pi}{4}\right) + (\pi f_3' t)^2 \times 5\cdot407\cos(0\cdot707\pi f_3' t - 56^\circ\,15') - \pi f_3' t \times 22\cdot83\cos(0\cdot707\pi f_3' t + 201^\circ\,47') + 49\cdot29\cos(0\cdot707\pi f_3' t - 70^\circ\,25')\right] \quad . \quad . \quad . \quad . \quad (5$$

All these formulae can be put into the form

$$v(t) = 1 + \varepsilon^{-0\cdot707\pi f_3' t}A(\pi f_3' t)\cos[0\cdot707\pi f_3' t + \phi(\pi f_3' t)] \quad . \quad . \quad . \quad . \quad (5$$

The decay time is obtained by finding a value of t for whic

$$\varepsilon^{-0\cdot707\pi f_3' t}A(\pi f_3' t) = \left.\begin{matrix}0\cdot01\\0\cdot001\end{matrix}\right\} \quad . \quad . \quad . \quad . \quad (58$$

and subtracting subsequently the corresponding value of t_l take from Reference 3.

(6.5) Over-Coupled Circuits

If n identical pairs of tuned circuits are considered coupled i such a way that the peaks of the amplitude response for eac pair are 1 dB above the amplitudes at the centre frequencies, the

$$\frac{f_3}{f_6} = \frac{0\cdot9076 + \sqrt{(0\cdot8237 + b)}}{0\cdot9076 + \sqrt{(0\cdot8237 + a)}} \quad . \quad . \quad . \quad (59$$

where a and b are defined by eqns. (23) and (24) respectively.

$n = 1$

$$v(t) = 1 - 1\cdot173\varepsilon^{-0\cdot522\pi f_3 t}\cos 0\cdot0852\pi f_3 t \quad . \quad (60$$

$n = 2$

$$v(t) = 1 + \varepsilon^{-0\cdot522\pi f_3' t}[\pi f_3' t \times 0\cdot688\cos(0\cdot852\pi f_3' t + 58^\circ\,30') + 1\cdot439\cos(0\cdot852\pi f_3' t + 1^\circ\,39')] \quad . \quad (61$$

$n = 3$

$$v(t) = 1 + \varepsilon^{-0\cdot522 f_3' t}[(\pi f_3' t)^2 \times 0\cdot2021\cos(0\cdot852\pi f_3' t - 31^\circ 15') + 1\cdot077\pi f_3' t\cos(0\cdot852\pi f_3' t + 47^\circ\,6') + 1\cdot779\cos(0\cdot852\pi f_3' t + 124^\circ\,12')] \quad . \quad (62$$

$n = 4$

$$v(t) = 1 + \varepsilon^{-0\cdot522\pi f_3' t}[-(\pi f_3' t)^3 \times 0\cdot0396\cos(0\cdot852\pi f_3' t + 58^\circ 30 + (\pi f_3' t)^2 \times 0\cdot3843\cos(0\cdot852\pi f_3' t - 41^\circ\,12') + \pi f_3' t \times 1\cdot48\cos(0\cdot852\pi f_3' t + 38^\circ\,42') - 2\cdot22\cos(0\cdot852\pi f_3' t - 63^\circ\,6') \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6$$

From the above formulae the decay time is found by a pro cedure analogous to that used in Section 6.4.

621.318.435.3.011.6

Monograph No. 140 M
July 1955

THE RESIDUAL TIME-CONSTANT OF SELF-SATURATING (AUTO-EXCITED) TRANSDUCTORS

By ULRIK KRABBE, Ph.D.

(*The paper was first received* 10*th December*, 1954, *and in revised form* 25*th March*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in July*, 1955.)

SUMMARY

It is well known that the time-constant of a self-saturating transductor is determined mainly by the control winding, increase in the resistance of which will have the effect of reducing the time-constant, but that the time-constant is also influenced by the main winding, so that it is not strictly proportional to the conductivity of the control circuit.

The paper deals with a theory for the effect of the main winding on the time-constant, and describes laboratory tests which confirm the theory; its consequences in cases where there is negative feedback from the output voltage are particularly discussed.

LIST OF SYMBOLS

f = Frequency of a.c. supply.
i_A = Instantaneous current in main winding of element A.
i_B = Instantaneous current in main winding of element B.
I_N = Average direct output current from the transductor.
I_S = Signal current.
I_1 = Primary current.
I_2 = Secondary current.
k_i = Current gain.
k_v = Voltage gain.
L = Inductance.
L_s = Inductance of control winding.
L_1 = Inductance of a primary winding.
L_2 = Inductance of a secondary winding.
m = Relative addition to time-constant from main winding.
M = Mutual inductance.
N_1 = Number of turns of a primary winding.
N_2 = Number of turns of a secondary winding.
p = Differential operator.
R_N = Load resistance.
R_V = Resistance of main winding.
R_i = Equivalent resistance of signal circuit.
R_s = Resistance of signal circuit.
v = Instantaneous supply voltage.
V_m = Average supply voltage.
V_N = Average direct output voltage.
V_s = Signal voltage.
α_0 = Firing angle.
$\beta_1, \beta_2, \beta_3, \beta_4$ = Angles of commutation.
β = Angle at which one self-saturation rectifier is conducting.
τ_s = Time-constant of control circuit (unsaturated).
τ_2 = Time-constant of secondary control winding.
τ_Y = Time-constant of feedback circuit.
τ_T = Time-constant of transductor.
τ_T' = Time-constant of transductor disregarding main-winding influence.
ω = Signal angular frequency.
ΔV_s = Actual control range of signal voltage.
ΔI_s = Actual control range of signal current.
$\dfrac{d\Phi}{d(NI)}$ = Slope of curve of flux/ampere-turns on unsaturated part of magnetization curve.

(1) DEFINITION OF TIME-CONSTANT

Several different definitions of time-constant are commonly found, and it is therefore necessary to state clearly which one is to be used. In practice, this means defining how the time-constant shall be measured.

For a rotating machine, especially a d.c. shunt generator, the following general expression gives the relation between input signal voltage, V_s, output voltage, V_N, and time:

$$V_N = k_v \frac{V_s}{1 + p\tau} \qquad (1)$$

where p is the differential operator, τ is the time-constant of the field (equal to L/R), and k_v is the static voltage gain.

A similar relation holds for a self-saturating transductor. In this case, however, the output consists of a number of pulses with a frequency equal to twice the supply frequency (for a single-phase transductor), and it is the average d.c. value of these which is of importance.

When the signal voltage (and therefore the control current) is changed, the output changes, as shown in Fig. 1. The most accurate means of measuring the average d.c. output in such a case is that obtained by plotting the average value of each pulse and joining the points with straight lines, but in this process there will be uncertainty as to the exact instant for which each point should be plotted. This means that the time-constant cannot be defined to an accuracy greater than a quarter of a cycle.

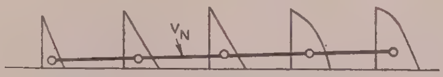


Fig. 1.—Output from a transductor when the control current is changed.

In some cases a time-constant is found by means of an oscillograph, but, again, since the instrument normally shows pulse peaks rather than averages, measurements of this kind are not very accurate.

Other difficulties in the measurement of the time-constant of a transductor are due, first, to the influence of size and direction of the input signal, both of which affect the blocking periods of the rectifiers unless the signal is very small, and secondly, to the non-linearity of the static characteristic, which means that the time-constant is different for different working ranges.

For all these reasons the frequency-response method of time-constant measurement, using very small input signals, has been greatly favoured. With this method it is possible to get much more accurate results than with, e.g. transient-response methods.

Correspondence on Monographs is invited for consideration with a view to publication.
Dr. Krabbe is with A/B Elektromekano, Hälsingborg, Sweden.

(2) THE FREQUENCY-RESPONSE METHOD APPLIED TO A SINGLE-STAGE AMPLIFIER

If a sinusoidal input is applied to an amplifier having the dynamic characteristic of eqn. (1), and the phase displacement between signal voltage and output voltage is plotted as a function of the logarithm of signal angular frequency ω, an arc-tangent curve of the form shown in Fig. 2 is obtained; when $\omega = 1/T$ there will be a 45° phase displacement.

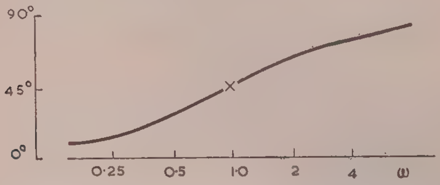


Fig. 2.—Phase-angle between output and input voltages plotted against signal angular frequency for a single-stage amplifier, e.g. an ideal self-saturating transductor or an ideal d.c. shunt generator.

Measurements with self-saturating transductors, when small inputs have been used and the phase displacement between the signal voltage and the component of the frequency ω in the output has been measured, have given actual curves very similar to Fig. 2. It is therefore reasonable to define the time-constant of such a transductor as follows:

> The time-constant of a self-saturating transductor is the reciprocal of that signal angular frequency which gives 45° phase displacement between a small sinusoidal signal voltage and the corresponding frequency in the output voltage, provided that the phase displacement follows an arc-tangent curve as a function of logarithm to signal frequency.

A time-constant is thereby defined for each operating point on the static characteristic.

During the experimental measurements leading to this definition, in order both to eliminate errors and to check the shape of the curve, the phase displacement for each determination of the time-constant was measured at different angular frequencies within a range of 0·5–64 rad/s, and a curve was drawn to determine the frequency which gave 45° displacement. The inverse of this frequency was regarded as the time-constant.

It should be noted that this method yields a value for the time-constant which has relatively little practical application when the speed of response to very large input signals is considered. It is nevertheless the only method whereby the theoretical relations between the time properties of the transductor and its other characteristics may be discovered.

(3) THEORY OF THE TIME-CONSTANT

The theory now generally accepted for the static and dynamic operation of a self-saturating transductor can be found in available literature,[1, 2] and will not be repeated in detail here. A short review may, however, be useful.

It may be shown that eqn. (1) applies to a transductor with self-saturation, and further, that in cases where the influence of the main winding on τ_T may be disregarded, the time-constant is given by

$$\tau'_T = \frac{k_v}{2f} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where k_v is the static voltage gain and f is the supply frequency; the prime (′) in τ'_T indicates that the influence of the main winding is neglected. This equation assumes the same number of turns on main and control windings, and in considering the output voltage neglects the internal voltage drop in the rectifiers and th main winding.

It is also known from elementary theoretical consideration that

$$\tau'_T = \frac{\tau_s}{2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3$$

where $\tau_s = \frac{L_s}{R_s}$, R_s being the resistance of the control circui L_s being equal to $N_s^2\frac{d\Phi}{d(NI)}$ and $\frac{d\Phi}{d(NI)}$ (for an ideal saturatio curve) being the slope of the unsaturated part of the dynami characteristic. In practice, the saturation curve is not ideal, an the slope $\frac{d\Phi}{d(NI)}$ is affected by eddy currents and hysteresis in complex manner; it can, however, be determined from the stati characteristic of the transductor.

If there is more than one closed circuit, e.g. a second signal c bias winding with N_2 turns and resistance R_2, then

$$\tau'_T = \frac{\tau_s}{2} + \frac{\tau_2}{2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4$$

where $\tau_2 = L_2/R_2$.

The main point, when considering the influence on the tim constant of the main winding, is that even when this winding short-circuited its influence is limited by the self-saturatio rectifiers, which are blocked for some parts of each cycle.

In some cases this influence is further reduced by the resistanc of the load on the winding, but in what follows the circuit c Fig. 3 only is considered, since it is known from theory an

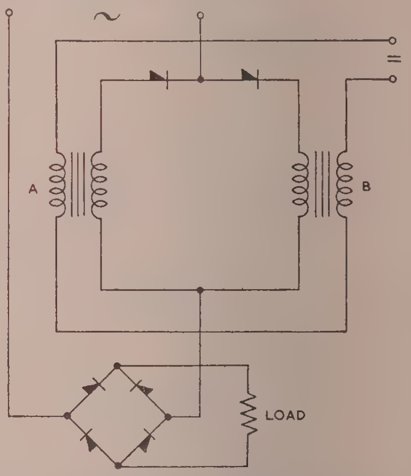


Fig. 3.—The self-saturating transductor discussed in the paper.

practice that the influence is largest in such a circuit and independent on the load resistance.

An expression for the influence of the main winding has bee given in Reference 1, but a more easily deduced relation is th given in Section 12.1. This relation is

$$\frac{1}{\tau_T} = 2\left(\frac{\pi - \beta}{\pi}\frac{1}{\tau_s} + \frac{\beta}{\pi}\frac{1}{\tau_s + \tau_V}\right) \quad . \quad . \quad . \quad ($$

where $(\pi - \beta)$ is the portion of each cycle during which there a blocking voltage across one of the self-saturation rectifie (e.g. A, Fig. 3), and τ_V is the time-constant of the main windi calculated in the same way as τ_s.

The correctness of this expression may be checked for $\beta = 18$ because τ_V is then added to the control-winding time-consta eqn. (4)], and it can also be seen that when $\beta = 0$ there can be no influence because one of the rectifiers will then always be blocked so that the main-winding circuit may be regarded as open.

In the elementary theory of the static operation of the transductor, where small excitation currents in the main windings are neglected, it is found that the rectifiers are blocked during each cycle between the firing angle α_0, where the one element saturates, and π. With this elementary theory we should thus find that $\beta = \alpha_0$.

Experiments show, however, that the addition to the time-constant from the main winding is less than that given by substituting α_0 for β in eqn. (5). The reason for this is that the self-saturation rectifiers are blocked for a longer period than that determined by the firing angle α_0. The increase is due to changes in small excitation currents flowing in the main winding of an element when it is not saturated, as will be explained in Section 4.

(4) BLOCKING PERIODS OF SELF-SATURATION RECTIFIERS

The following analysis relates to a transductor of the type shown in Fig. 3, with a core characteristic as shown in Fig. 4. A purely resistive load and an alternating input waveform of rectangular shape are assumed.

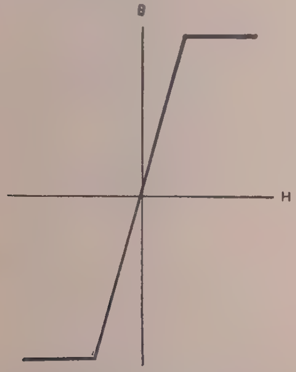


Fig. 4.—Idealized saturation curve for the transductor cores.

(4.1) Case in which the Control-Circuit Resistance is Infinite or Completely Smoothed by a Reactor

Fig. 5(*a*) shows the alternating voltage and corresponding current in the two transductor elements, and Fig. 5(*b*) shows how the voltage is divided between induced voltage in the element, voltage drop in the load and winding resistance, and blocking voltage in the rectifiers.

Consider element A when $\alpha = 0$ and the magnetic induction is just beginning to rise. The instantaneous voltage across the element is

$$v + i_B R_N - i_A R_N$$

where v is the alternating voltage and i_A and i_B are the small currents in the elements. This voltage is divided between induced voltage and voltage drop $i_A R_V$.

At $\alpha = \alpha_0$, the element becomes saturated, and i_A increases suddenly to a value limited only by $R_N + R_V$; then when $\alpha = \pi$ the alternating voltage changes direction and the induced voltage in element A reverses its direction. This voltage is $v - i_B R_N + i_A(R_N + R_V)$ and it therefore regains its original state very shortly before $\alpha_0 + \pi$. Between this point and $\alpha_0 + \pi$ there will be, as can be seen from the diagram, a very high blocking voltage across the self-saturation rectifier; this will begin to drop at $\alpha_0 + \pi$.

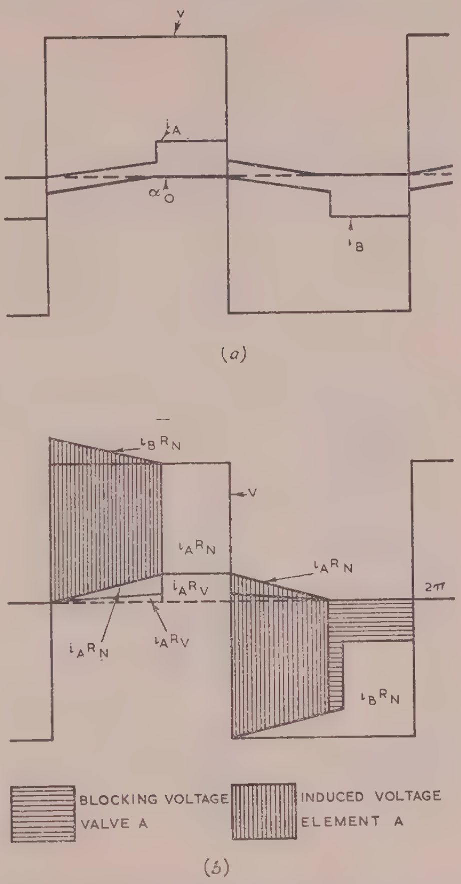


Fig. 5.—Action of a transductor with a saturation curve as shown in Fig. 4, having fully smoothed control current (or no control current).

It is clear that the high blocking voltage is a consequence of the change, due to voltage drop in R_V, of the small excitation current in the main winding. In practice, the peak will be reduced because of the actual shape of the saturation curves of the iron cores, which differ from the ideal.

(4.2) Case in which the Control-Circuit Resistance is Zero

Even though the condition of zero resistance cannot be realized in practice, it is of great interest. If the control-circuit resistance were zero, the induced voltage in the two elements at every instant would be exactly equal but opposite in direction. Thus, if a current i_A were flowing in element A this would give a voltage drop $i_A R_V$, and a corresponding blocking voltage would appear across rectifier B.

Hence, only one element could be carrying current at any one time and each rectifier would be blocked for exactly one half of every cycle. This would be possible because there would be alternating current in the control winding.

The mode of operation is shown in Fig. 6, in which (*a*) shows the alternating voltage and the two element currents; (*b*) shows how the voltage is divided between induced voltage, resistive voltage drop, and blocking voltage; and (*c*) shows the current in the control winding.

At $\alpha = 0$ the magnetic induction in element A is starting to rise. This cannot be due to current in the main winding, but is

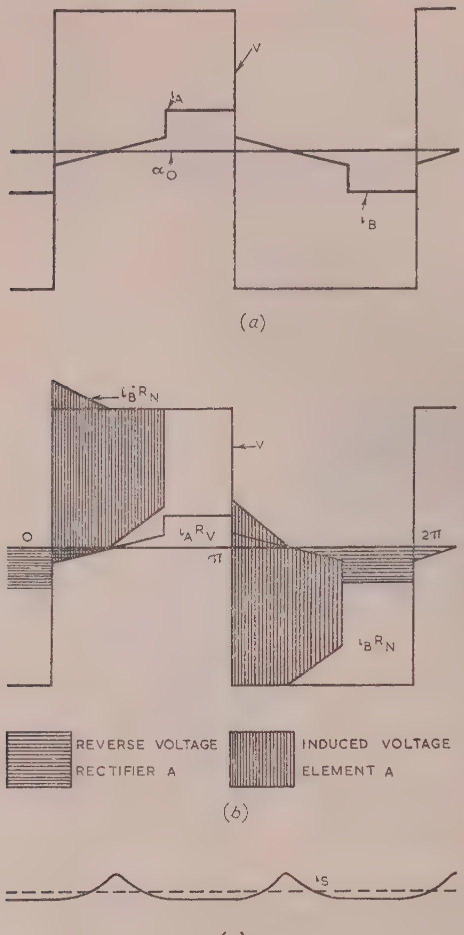


Fig. 6.—Action of a transductor corresponding to Fig. 5 but with fully short-circuited control windings.

caused by an increasing current flowing in the short-circuited control winding. At $\alpha = \alpha_0/2$ the main winding begins to carry current, with twice the change corresponding to the flux change. However, at the same time, the current in the control winding is decreasing, so that the difference gives the necessary excitation current. The diagrams show that the induced voltages in the two elements are thus equal.

(4.3) Comparison of Cases in Sections 4.1 and 4.2

It is interesting to compare the two cases of infinite and zero control-circuit resistance. In the first case the blocking angle is $\pi - \alpha$ and in the second it is π, and for all practical values of control-circuit resistance the angle must therefore lie between these limits. It is possible to confirm this by direct calculation, but the values so obtained are unlikely to be accurate because of various unavoidable approximations.

With sinusoidal alternating voltages the results are almost the same as with rectangular waveforms.

(5) EXPERIMENTAL MEASUREMENT OF BLOCKING ANGLES

The transductor used for experimental measurements had the characteristics given in Section 12.2. It will be seen that these characteristics were made as nearly ideal as possible.

To determine the blocking voltages across the rectifiers it was found necessary to make a special battery-driven d.c. electronic amplifier, since the measurements would otherwise have been upset by earthing difficulties in connection with the capacitances of rectifiers and cores.

Figs. 7 and 8 show oscillograms of the blocking voltage across the self-saturation rectifiers. Fig. 7 relates to an open control circuit and Fig. 8 to a short-circuited control winding. Compared with the diagrams of Figs. 5 and 6, there is good agreement even if the peak of the blocking voltage for Fig. 7 is reduced by the non-ideal properties of the core material.

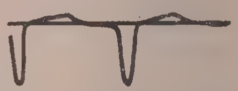


Fig. 7.—Voltage waveform across the self-saturation rectifier described in Section 12.2 with an open control winding.

The waveform corresponds to the theoretical relations represented in Fig. 5.

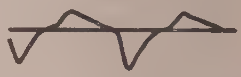


Fig. 8.—Waveform as for Fig. 7 but with short-circuited control winding, corresponding to the theoretical relations in Fig. 6.

The scale for Fig. 7 has been reduced to a little less than half that of Fig. 8.

Measurement of the blocking angles (i.e. the intervals in each cycle during which the rectifiers were blocked) was made as follows.

A cathode-ray oscillograph was used, with the output from the electronic amplifier providing X-deflection and a sinusoid of controllable phase the Y-deflection. The phase regulator was

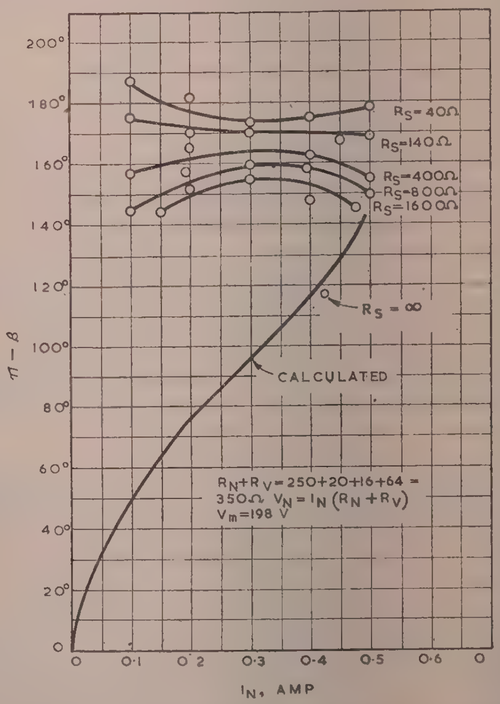


Fig. 9.—Measured blocking angles for the transductor as a function of output voltage (current) for different values of resistance in the control circuit.

first set so that the start of the blocking interval was at a certain point on the screen and then turned until the end of the interval was at the same point, the difference between the two readings on the regulator thus being the angle of the blocking interval. Readings could be made to an accuracy of 1 or 2°.

Fig. 9 shows blocking angle as a function of output current (or output voltage) for different values of resistance in the control circuit. It can be seen from this Figure that the angle is nearly equal to π even when the control-circuit resistance is high. The calculated value is $\pi - \alpha_0$, from the relation between alternating voltage and output voltage. The experiments showed that there was no relation between blocking angle and firing angle.

The experimental results for the open-control-circuit condition is close to the calculated value of $\pi - \alpha_0$ (see Fig. 5).

It is apparent that angles greater than π can occur. No theoretical explanation for this has yet been found, although it cannot be due to errors in measurement.

It is worth mentioning that at both ends of the blocking interval the blocking voltage on the rectifiers was extremely small, so that the blocking angle could easily be changed by even a small transient signal. This effect would probably be less pronounced with other self-saturating circuits without separate rectifiers for self-saturation.

(6) THE CONSEQUENCES OF EQN. (5)

A common requirement is to reduce the time-constant of a transductor without too great a reduction in amplification. When the resistance of the control winding is increased, the time-constant will be reduced, but, according to eqn. (5), not as much as the amplification.

If m denotes the relative addition contributed by the main winding to the time-constant as determined by the control winding only, we have

$$m = \frac{\tau_T - \tau_T'}{\tau_T'} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

and eqns. (6) and (5) will give

$$m = \frac{\dfrac{\beta}{\pi}}{\dfrac{\tau_s}{\tau_V} + \dfrac{\pi - \beta}{\pi}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

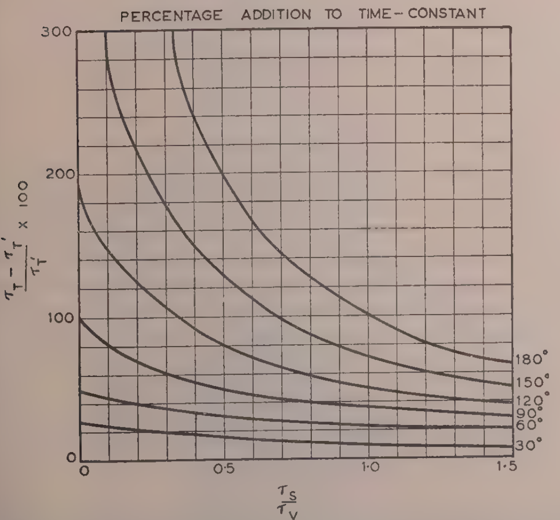


Fig. 10.—Percentage addition to the calculated time-constant τ_T' from the main winding as a function of τ_S/τ_V for different values of β (blocking angle $= \pi - \beta$).

This relation shows that, theoretically, the value of τ_T can always be made as small as desired; it is shown graphically in Fig. 10. Even if β is a function of R_S, the addition m will never exceed a certain value (corresponding to $\beta = \alpha_0$). This means that the addition to the time-constant τ_T' is dependent on τ_T' itself. It also varies with R_S, partly as a result of the simple relation between τ_S and R_S [which may be incorporated in eqn. (7)], and partly because of the more complicated relation between R_S and the blocking angle, shown as measured in Fig. 9.

It should be noted in passing that the expression "residual time-constant" normally refers only to the half-cycle delay resulting from the nature of the output, while the addition m just discussed has no relation to the supply frequency.

(7) MEASUREMENT OF TIME-CONSTANTS

Experimental measurements of the time-constant of the transductor described in Section 12.2 were made with the circuit shown in Fig. 11. The variable-frequency generator, which had a frequency range of 0·5–64 rad/s, consisted of a rotating potentiometer with sinusoidally-distributed resistance connected to a commutator with 100 segments. The winding was fed with direct current from slip rings as shown. One alternating voltage was taken from two fixed brushes, and a second voltage was tapped from a movable brush on the commutator and a slip ring connected to the centre of the potentiometer. The phase displacement between the two voltages, which could thus be varied, was read on a scale showing the position of the movable brush.

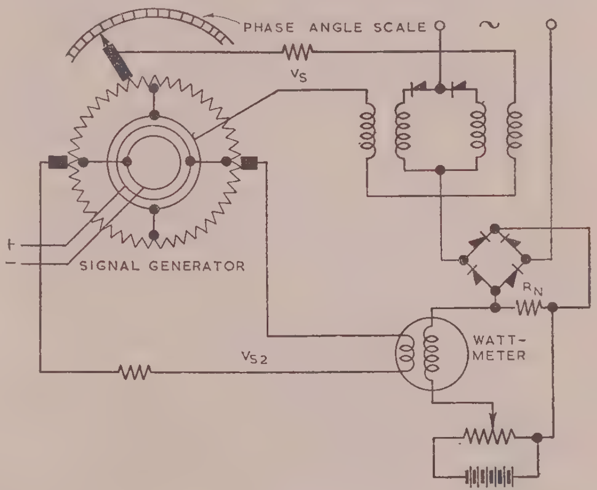


Fig. 11.—Arrangement for the measurement of the phase-shift between input voltage V_S and output voltage V_N of a transductor.

The phase-shift between input voltage and output voltage was determined by using a wattmeter as a zero-reading instrument, the movable brush being adjusted until the wattmeter gave no reading. The d.c. output from the transductor was compensated by a battery, as shown in the Figure, to reduce the movement of the wattmeter pointer. This method enabled readings to be made to an accuracy of 1 or 2°.

At low values of the frequency ω, the wattmeter had too little damping and it was necessary to use instead an oscillograph in conjunction with a quick-acting filter for the pulses in the output voltage. This method of measurement was rather less accurate than the wattmeter method and larger signals had to be used. With the wattmeter the output varied about 5 volts; with the oscillograph the variation was about 15 volts.

To avoid influence from bias windings, all measurements were made at the point in the static transductor characteristic where I_S was zero.

(7.1) Direct Determination of the Time-Constant

The direct process, which was used to check the results of the previous determination, consisted in measuring τ'_T and τ_T with R_S as a parameter. τ'_T was determined from the slope of the static characteristic at the point where $V_S = 0$, and since this characteristic changes its shape slightly when R_S is changed, τ'_T was measured independently for each value of R_S. τ_T was measured as already described by means of the phase-shift between V_S and V_N.

Figs. 12 and 13 show the results of the measurements, the full line and circle points in Fig. 12 giving $\tau_T - \tau'_T$ and those in Fig. 13 giving $m = \frac{\tau_T - \tau'_T}{\tau'_T}$.

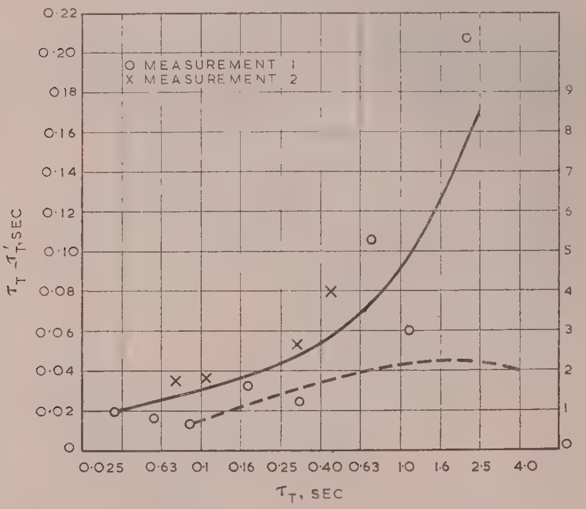


Fig. 12.—Comparison between measured values of $\tau_T - \tau'_T$ (full-line curve) and calculated value (dotted curve) as a function of the measured time-constant τ_T.

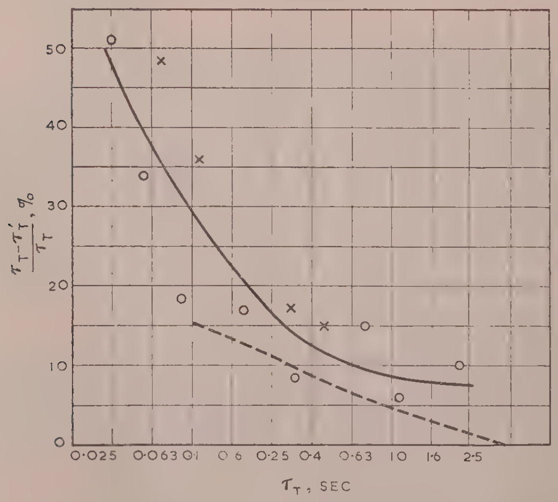


Fig. 13.—Measured and calculated values of $\frac{\tau_T - \tau'_T}{\tau_T}$.

The dotted curves were calculated in the following way: The values from Fig. 9 were inserted in eqn. (5) to determine β, and τ_S was determined as $2\tau'_T$ from the measurement abov

$\tau_V = \tau_S \frac{R_S}{2R_V}$.

The agreement between the measured and calculated value (on the basis of measured blocking angles) is seen to be quit satisfactory. The measured addition is a little higher than th calculated value, but the difference in the range with the bes time-constant determination is within 0·5–1·0 cycle and there i reason to expect a delay of a little less than half a cycle. This i because a signal in the beginning of the half-cycle will first b represented in the output in the later part of the same half-cycl and a signal in the later part of a half-cycle will first be represented in the next half-cycle.

The cross-points on Figs. 12 and 13 show the results of directl measuring the addition m from the main winding, as described i Section 7.2.

(7.2) Measurement of the Addition to the Time-Constant

Fig. 14 shows a magnetic core on which the primary windin corresponds to the control winding and the secondary windin corresponds to an additional time-constant. The followin relations apply:

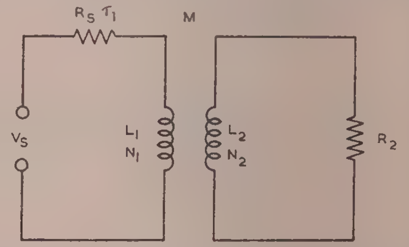


Fig. 14.—Diagram of magnetic core with an excitation winding an a short-circuited secondary winding, to illustrate measurement o additional time-constant.

$$(R_S + pL_I)I_1 + pMI_2 = V_S \quad . \quad . \quad . \quad . \quad (8$$

$$pMI_1 + (R_2 + pL_2)I_2 = 0 \quad . \quad . \quad . \quad . \quad (9$$

$$I_7 = \frac{V_S}{R_S}\frac{1 + p\tau_2}{1 + p\tau_S + p\tau_2} \quad . \quad . \quad . \quad . \quad (10$$

and

$$I_1N_1 + I_2N_2 = \frac{V_SN_1}{R_S[1 + p(\tau_S + \tau_2)]} \quad . \quad . \quad . \quad (11$$

$$\frac{I_1}{I_1N_1 + I_2N_2} = \frac{1 + p\tau_2}{N_1} \quad . \quad . \quad . \quad . \quad (12$$

This means that the phase-shift between the primary curren and the field in the core is dependent on the secondary time constant only. For the transductor the output voltage follow the mean flux and the equation shows that the phase-displacemen between control current, I_S, and output voltage V_N shoul depend exclusively on the addition from the main winding.

The additional time-constant was measured in a similar wa to τ_T, but in this case from the phase-shift between I_S and V_l for different values of R_S, and with the difference that a therma wattmeter coupling was used, to avoid any influence on th control current.

Considering how small in comparison with the supply frequency were the time-constants measured in this way, the agreement between the indirect and direct measurements seems t confirm that eqn. (5) is useful. With large time-constants, wher the oscillograph method was used to determine phase-shift, th agreement was not so good, and one reason for this may b that it was necessary to use larger signals, which influenced the blocking interval. The oscillograms also showed traces of even harmonics of the signal frequency in the output voltage, and this might be explained similarly by the signal size.

(8) EFFECTS OF COUPLINGS WITH NEGATIVE FEEDBACK FROM THE OUTPUT VOLTAGE

Much consideration has been given in recent years to increasing the speed of transductors by the use of couplings with negative feedback from the output voltage. To simplify the analysis of this method it is convenient to assume the same number of turns in the control winding as in the main winding.

It is shown in Reference 1 that the relation

$$\tau'_T = \frac{1}{2f}\frac{dV_N}{dV_S}$$

holds equally with positive and negative feedback (without delay in the feedback circuit), and also that the percentage addition of a secondary time-constant to the time-constant from the control winding is not altered when the total time-constant τ_T is changed by means of feedback. In fact, this is a consequence of energy-storing in the transductor.

Consider now a transductor coupling, with equal numbers of turns in the two windings, having a current gain k_i, a voltage gain k_v, and a time-constant (neglecting the addition from the main winding) $\tau'_T = k_v/2f$.

Assume also that the characteristic includes some complex non-linearity, and examine the effects of changing the coupling circuit in three different ways, as shown in Figs. 15, 16 and 17.

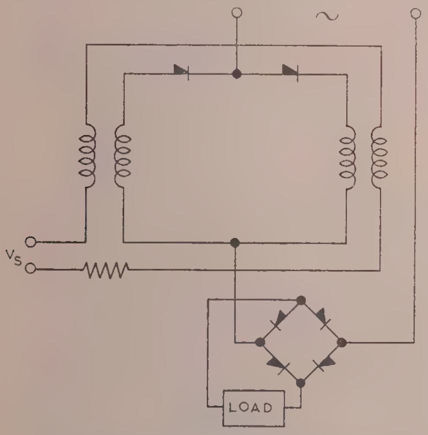


Fig. 15.—Transductor coupling in which the time-constant is reduced by means of a series resistance.

The arrangement of Fig. 15 is designed to reduce the time-constant by means of a resistor in series with the control winding so that $k_v = 1$. Fig. 16 has negative feedback from the output voltage (by summation of voltage), while Fig. 17 provides negative feedback from the output voltage by means of a second control winding (by summation of magnetomotive forces) and is assumed to be so arranged that $k_v = 1$.

The properties of the three couplings are compared in Table 1.

It can be seen that the main advantage of the arrangement shown in Fig. 16 is increased linearity compared with that of Fig. 15. It is obvious that the coupling in Fig. 17 has no practical use with so high a degree of feedback.

Consider now the effects, on the circuit of Figs. 15 and 16, of an additional time-constant due to the main winding.

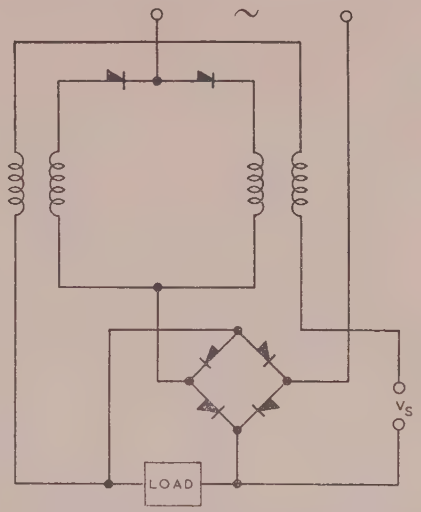


Fig. 16.—Transductor coupling in which the time-constant is reduced by means of negative feedback.

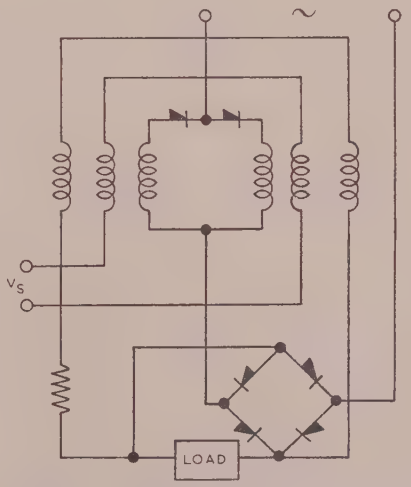


Fig. 17.—Transductor coupling in which the time-constant is reduced by means of negative feedback from a second control winding.

Table 1

	Power gain	Time-constant	Linearity
Coupling as in Fig. 15 ..	k_i	$1/2f$	—
Coupling as in Fig. 16 ..	k_i	$1/2f$	Increased
Coupling as in Fig. 17 ..	k_i/k_V	$(1/2f)(\tau_S+\tau_Y)/\tau_S$	Increased

τ_Y is the time-constant of the secondary control circuit shown in Fig. 17.

For Fig. 15 it has already been shown that the time-constant $1/2f$ must be multiplied by a certain factor. The arrangement in Fig. 16 is affected quite differently. An additional time-constant alters the phase relation between control current and output voltage, so that when, for example, the addition is due to an increase in V_S, its first result will be a control current greater than that in the stationary state. It can also be said that the output power through the control winding will momentarily come from the input power source, because the impedance against a surge impulse of the control winding is reduced by the secondary time-constant. The result of this will be not an addition to the time-constant from the main winding, but an overshoot of the control current during transient conditions.

Normally, the input power is small, which is the reason for the use of the amplifier. The input power source may therefore be considered as an e.m.f. in series with an impedance, and with impedance matching this will be of magnitude $R_i = \Delta V_S/\Delta I_S$, where ΔI_S and ΔV_S are the variations in the range over which the amplifier is used.

The overshoot in control current during transient conditions will, however, give a voltage drop in the series resistance R_i, and thus will reduce the signal, so delaying the transductor. It is therefore almost as important to keep the additional time-constant low for the circuit of Fig. 16 as for that of Fig. 15.

What has been said about the additional time-constant from the power winding applies also for time constants from bias windings.

The circuit for negative feedback where the control winding itself is used to produce a voltage equal to the output voltage, by Ramey,[5] differs not in this respect from Fig. 16, except that the rectifiers will limit the overshoot (dependent on magnitude and direction of the signal), so that a time delay will occur partly as for Fig. 15.

It has been suggested[4] that the negative feedback circuit of Fig. 16 should be especially advantageous for transductors using magnetic material with a rectangular hysteresis loop. In this connection it is worth remembering that so high a degree of negative feedback is not normally necessary and that the couplings of Figs. 15 and 16, or 15 and 17, can be combined so that just sufficient feedback is introduced to give a suitable linearity, while the time-constant is reduced by a resistance in series with the control winding.

A great advantage, however, of the coupling in Fig. 16 is that a signal in the negative direction will not increase the output, a property not possessed by most transductor couplings.

(9) CONCLUSIONS

The theoretical and experimental evidence given here shows that the main winding of a self-saturating transductor influences the time-constant for small signals to an extent dependent on the blocking intervals of the self-saturation rectifiers and the resistance of the main winding. In addition, however, the blocking interval is evidently directly dependent on the amplitude of the output signal; this agrees with the normal experience that a transductor usually has a faster response when its output is increased, this effect being dependent on the ratio of signal amplitude to blocking voltage.

When making a fast transductor it is therefore important to choose a circuit where the blocking voltage across the self-saturation rectifiers is large, and this is why the circuits without separate self-saturation rectifiers are, as is generally known, better than that shown in Fig. 3. An additional reason for this may be the greater number of rectifier plates, which at small currents will reduce the time-constant of the main winding considerably because the resistance is larger.

When selenium rectifiers are used a further reason is that a large signal for increasing output will increase the size of the blocking voltage and thus increase the leakage current. This phenomenon will be less pronounced for transductors without separate self-saturation rectifiers than for that of Fig. 3, given the same value of blocking voltage for the rectifiers in the two cases.

(10) ACKNOWLEDGMENTS

The author wishes to thank Thomas B. Thrige, Odense, Denmark, for permission to publish the paper, and Dr. E. H. Frost-Smith for help in editing it.

(11) REFERENCES

(1) KRABBE, U.: "The Transductor Amplifier" (Munksgaard, Copenhagen, 1949).

(2) MILNES, A. G., and LAW, T. S.: "Auto-Self-Excited Transductors and Push-Pull Circuit Theory," *Proceedings I.E.E.*, Paper No. 1599 M, April, 1954 (**101**, Part II, p. 643).

(3) LAMM, U.: "The Transductor" (Esselte Aktiebolag, Stockholm, 1943).

(4) SCORGIE, D. G.: "Fast Response with Magnetic Amplifiers," *Transactions of the American I.E.E.*, 1953, **72**, Part I, p. 741.

(5) RAMEY, R. A.: "On the Mechanics of Magnetic Amplifier Operation," *ibid.*, 1951, **70**, Part II, p. 1214.

(6) DUNNEGAN, T., and HARNDEN, J. D.: "The Cyclic Integrator. A Device for Measuring the Frequency Response of Magnetic Amplifiers," *ibid.*, 1954, **73**, Part I, p. 358.

(12) APPENDICES

(12.1) Deduction of a Formula for the Influence of the Main Winding on the Time-Constant of a Transductor

The formula derived below is given in Reference 1 but with different assumptions.

Consider the circuit of Fig. 3, which, neglecting the leakage reactance between the main winding and the control windings, is equivalent to a circuit proposed by M. O. Jorgensen of Copenhagen and shown in Fig. 18. The time-constant is found by determining the change of flux in one element in half a cycle when a voltage V_S is applied to the control winding.

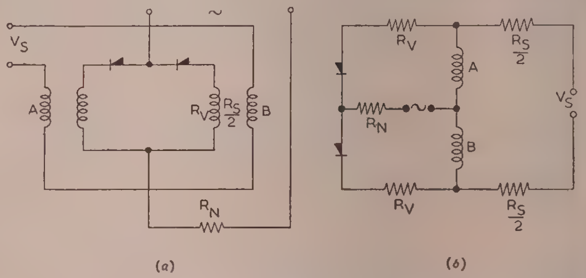


Fig. 18.—(*a*) Transductor circuit discussed in the paper. (*b*) Equivalent diagram assuming no leakage field between control winding and main winding.

The resistance of the control winding is R_S and the self-inductance of each element is L when unsaturated and 0 when saturated. τ_S for the control circuit is $2L/R_S$.

It is clear that four different cases must be examined: when the rectifiers are open (conducting) and blocked, and also when the elements are saturated and unsaturated. For each interval the law of superposition may be used.

An expression for the change of output voltage in half a cycle may be found from the change of flux in half a cycle for one element due to the control voltage V_S.

Taking first an interval where the flux in A is decreasing and that in B is increasing, if at such a time a control voltage V_S is applied in the direction giving increased output, this will cause the flux in A to drop more slowly and the flux in B to rise faster, reaching saturation a short time before the stationary firing angle α_0. Element A will therefore "inherit" flux from element B. This phenomenon is described in References 1 and 3.

The consequence is that the total change of flux for one element is equal to the sum of the changes from the stationary state for both elements in each part of the half-cycle.

Fig. 19 shows the voltage v and the element currents i_A and i_B, representing a general case in which is covered both Fig. 5 and

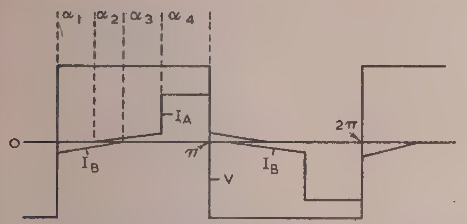


Fig. 19.—General relation between alternating voltage and the two currents in main windings of the transductor, corresponding to the two cases in Figs. 5 and 6.

Fig. 6, no assumption having been made concerning the lengths of the intervals α_1, α_2, α_3 and α_4.

Table 2 shows the conditions of the rectifiers and elements in the four intervals.

Table 2

Interval	Element		Rectifier	
	A	B	A	B
α_1	Unsaturated	Unsaturated	Blocked	Open
α_2	Unsaturated	Unsaturated	Open	Open
α_3	Unsaturated	Unsaturated	Open	Blocked
α_4	Saturated	Unsaturated	Open	Blocked

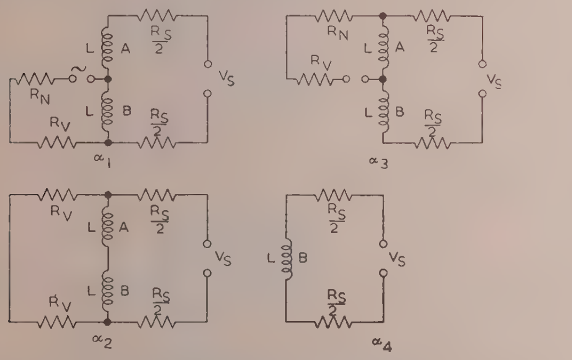


Fig. 20.—Simplified equivalent diagrams corresponding to Fig. 18 for the four intervals α_1, α_2, α_3 and α_4.

For each of these four cases an equivalent diagram may be drawn, as shown in Fig. 20, and the products of voltage and interval-length for the two transductor cores given in Table 3 may then be found.

Table 3

Interval	Product of voltage and interval-length	
	Element A	Element B
1	$V_S\alpha_1$	0
2	$\frac{V_S}{2}\frac{R_V}{R_V+R_S/2}\alpha_2$	$\frac{V_S}{2}\frac{R_V}{R_V+R_S/2}\alpha_2$
3	0	$V_S\alpha_3$
4	0	$V_S\alpha_4$

The sum of these terms is inversely proportional to the time-constant, thus

$$\frac{1}{\tau_T} = k_1 V_S\left(\alpha_1 + \alpha_3 + \alpha_4 + \alpha_2\frac{R_V}{R_V + R_S/2}\right) \quad . \quad (13)$$

If the influence from the main winding were neglected, this would become

$$\frac{1}{\tau_T} = k_1 V_S \pi = \frac{2}{\tau_S} \quad . \quad . \quad . \quad . \quad . \quad (14)$$

which gives

$$\frac{1}{\tau_T} = \frac{2}{\tau_S}\left(\frac{\alpha_1 + \alpha_3 + \alpha_4}{\pi} + \frac{\alpha_2}{\pi}\frac{R_V}{R_V + R_S/2}\right) \quad . \quad . \quad (15)$$

Now, if the blocking interval for the self-saturation rectifier is designated $\pi - \beta$, introducing $\tau_V = \frac{L_V}{R_V}$ gives

$$\frac{1}{\tau_T} = 2\left(\frac{\pi - \beta}{\pi}\frac{1}{\tau_S} + \frac{\beta}{\pi}\frac{1}{\tau_S + \tau_V}\right) \quad . \quad . \quad . \quad (16)$$

This expression is given in Reference 1, except that there the firing angle is α_0 instead of β as a consequence of too simplified assumptions. Similar expressions for other transductor couplings given in Reference 1 may therefore be assumed to hold when α_0 is replaced by β, where $\pi - \beta$ is the real blocking angle for the rectifiers.

It should be noted that the time-constant τ_S in the expressions cannot be accurately deduced from the slope of the saturation curve, because the small excitation current in the main winding will change the slope of the static characteristic of the transductor from that of this curve. When using eqn. (16) it is therefore necessary to determine τ_S from the measured voltage-gain on the static characteristic for each point, thus

$$\tau_S = \frac{dV_N}{dV_S}\frac{1}{f} \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

with the same number of turns in the windings and with the internal voltage drop included in V_N.

(12.2) Details of the Transductor used in the Experiments

(12.2.1) Cores

The transductor used for the experiments was made from interleaved stampings with the dimensions shown in Fig. 21.

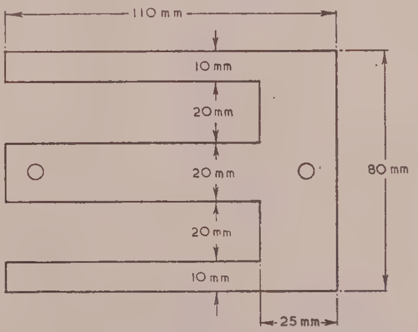


Fig. 21.—Shape and dimensions of the laminations used.

The stampings were of cold-rolled steel, Armco Trancor XX grade, annealed in a protective atmosphere after stamping. Each core consisted of 72 layers with a total stack height of 30 mm. The rolling direction was parallel to the legs and the average length of the magnetic path was 23 cm. The frequency of the supply was 50 c/s.

Each core had two windings of 0·45 mm-diameter double-enamelled wire, each winding being of 1 250 turns and the two being wound as a double wire, one for the control winding and the other for the main winding. This method involved some risk of insulation breakdown, but ensured equal numbers of turns and eliminated leakage reactance between the two windings. The capacitance between the windings was found to be 0·04 μF, corresponding to 80 000 ohms at 50 c/s.

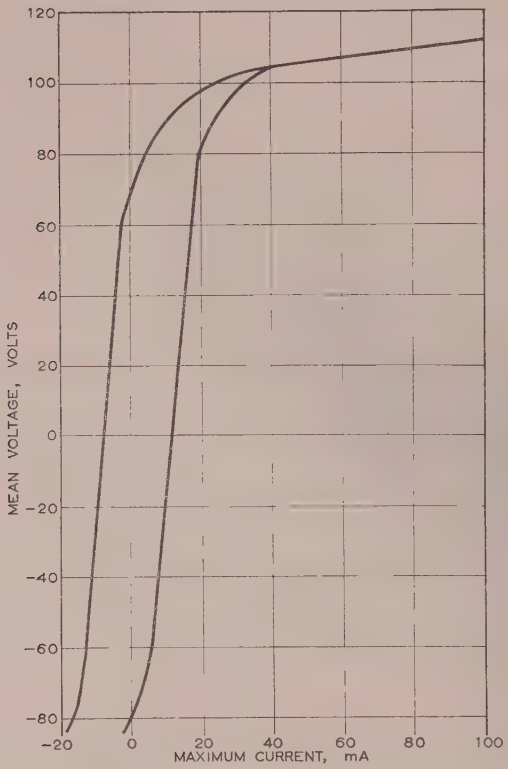


Fig. 22.—Hysteresis loop for one transductor element, measured at 50 c/s.

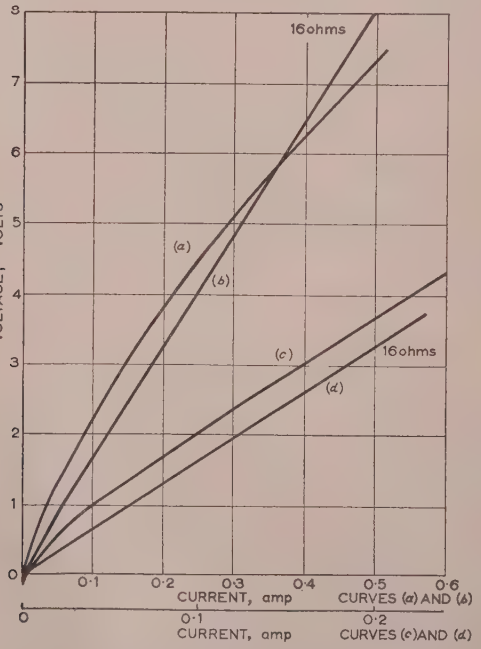


Fig. 23.—Relation between direct voltage-drop and current for the rectifier valves (8 anodes in parallel).

The resistance of each winding was 20·7 ohms at 20° C. The copper in each core, when fed with direct current, could dissipate 14 watts for a rise in temperature of 60° C.

Fig. 22 shows the B/H curve for the core, measured at 50 c/s given in terms of the mean voltage induced in the main winding at the particular value of B, plotted against maximum current measured by the method described in Reference 1, using a synchronous operating contact in connection with a d.c. voltmeter.

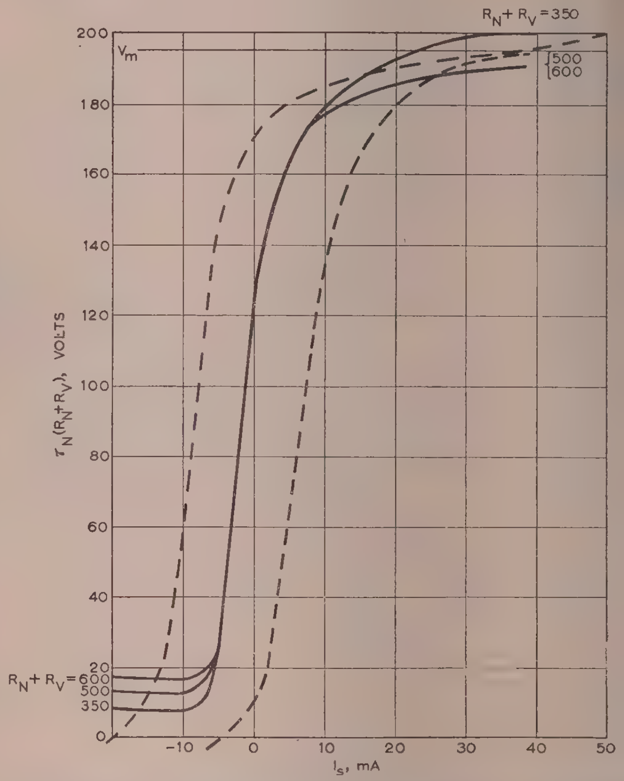


Fig. 24.—Output voltage $[I_N(R_V+R_N)]$ as a function of control current for different values of $R_V + R_N$.

(12.2.2) **Rectifiers.**

To secure characteristics as linear as possible and to avoid leakage currents, hot-cathode vacuum tubes were used (Philips type Cy. 2). For each of the two self-saturation rectifiers, four tubes each with two anodes were connected in parallel. The characteristic is shown in Fig. 23. The resistance of each rectifier was 16 ohms. The resistance of 16 ohms for the self-saturation rectifier was included in R_V, so that R_V was 36 ohms.

The main rectifier was bridge-connected, with four anodes in parallel for each branch. In all, 16 tubes were used in the circuit.

The resistance of the main rectifier (64 ohms) was included in R_N.

All measurements of time-constants were made with a load resistance of 250 ohms, and the static characteristics were calculated so that $V_N = I_N(16 + 20 + 250 + 64) = I_N \times 350$ ohms. It is obvious that the non-linear characteristic of the rectifier gives an almost negligible error in this determination.

Fig. 24 shows static characteristics at three different load resistances with $V_m = 220 \times 0{\cdot}9 = 198$ volts. On this graph the dotted line is the saturation curve from Fig. 22 displaced so that the point at 50 mA gives 198 volts.

There is satisfactory agreement between the measured and calculated curves, considering that hysteresis influences the characteristic in a complicated manner.

621.318.435.3:621.316.722

Monograph No. 141 M
Aug. 1955

A.C. CONTROLLED TRANSDUCTORS

By A. G. MILNES, M.Sc., Associate Member, and T. S. LAW, B.Sc.

(*The paper was first received* 29*th December*, 1954, *and in revised form* 31*st March*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in August*, 1955.)

SUMMARY

An analysis of the behaviour of the single-core, auto-self-excited, transductor element is made for the condition when the control circuit has finite resistance. Consideration is given to control by direct-voltage signal as an introduction to the performance with an alternating-voltage signal of the same frequency as the supply. Factors influencing the output characteristic are examined, and methods of improvement, such as phase shift and special bias arrangements, are discussed.

The a.c. control of full-wave transductors and some push-pull circuits with half-wave and full-wave outputs are referred to. The behaviour of a typical push-pull 4-element design is examined experimentally, and its performance when the control is by alternating-voltage signal is compared with that for direct-voltage control.

LIST OF PRINCIPAL SYMBOLS

i = Instantaneous current in the main circuit, amp.
i_c = Instantaneous current in the control circuit, amp.
I_c = Control current (mean), amp.
L = Inductance of a single-element main winding, H.
L_b = Inductance of a single-element bias winding, H.
L_c = Inductance of a single-element control winding, H.
M = Mutual inductance of the control and main circuit, H. $M = \sqrt{LL_c}$.
N_a = Number of turns in the main winding of each transductor element.
N_b = Number of turns of bias winding on each element.
N_c = Number of control turns on each element.
N_f = Number of feedback turns on each element.
R = Load resistance, ohms.
R_b = Total resistance of the bias circuit, ohms.
R_c = Total resistance of the control circuit, ohms.
R_f = Total resistance of feedback circuit, ohms.
R_s = Series-summing resistance, ohms.
V_{av} = Average load voltage, volts.
V_c = Signal voltage applied to the control circuit, direct or peak value, volts.
V_p = Peak transductor supply voltage, volts.
V_r = Reverse voltage sustained by the rectifier, volts.
α = Angle at which the transductor element saturates.
β = Angle at which the rectifier stops conducting.
γ = Angle at which the rectifier starts conducting.
η = arc tan $(\omega L_c/R_c)$.
ψ = Angular displacement of signal voltage relative to supply voltage.
Φ_s = Saturation flux per single core, maxwells.
ω = Angular frequency of supply, rad/sec.

(1) INTRODUCTION

Transductor theory has been developed fully for d.c. control conditions with the signal windings arranged on two cores in series opposition so that the voltages induced in the control circuit are not of large resultant magnitude.[1] For these circuit conditions the transductors respond to d.c. and low-frequency signals. For signals of the same frequency as that of the transductor supply source, series-opposition control windings are not used, but instead the control circuit contains a high impedance to restrict the effect of voltage induced from the supply circuit. This class of circuit with a.c. control has characteristics rather different from those of normal d.c.-controlled transductors, but very little information on this matter is available in the literature of the subject.

A single-core transductor with a main winding, N_a, in series with an auto-self-excitation rectifier and a resistance R representing the load, and with a control winding N_c connected to a source of signal through a large resistance R_c, is shown in Fig. 1(*a*).

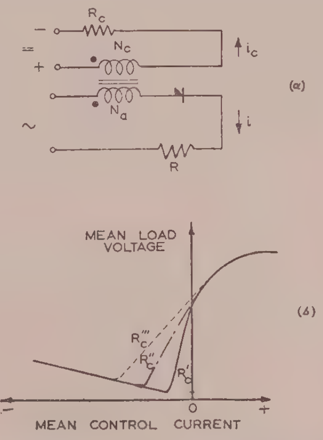


Fig. 1.—Single-element auto-self-excited transductor.

(*a*) Circuit with direct-voltage control.
(*b*) Typical characteristics with change of control-circuit resistance.
$R_c''' < R_c'' < R_c'$

To provide a suitable foundation for the subsequent development of a.c. control theory, it is desirable first to consider the behaviour of this circuit with a direct-voltage signal applied to the control circuit. Previous explanations of the action of this circuit have been over-simplified or idealized by the assumption that the control current is completely smoothed by a large choke in the circuit. An exception is the treatment by E. J. Smith, who shows that with finite resistance, R_c, control characteristics as in Fig. 1(*b*) are obtained. From these, reduction of the control-source resistance is seen to decrease the sensitivity of the region of the characteristic normally used.[2] Characteristics similar to those of Fig. 1(*b*) are derived analytically in Section 2, and the idealized waveforms for the action are presented for a full understanding of the circuit behaviour. With this background the changes which occur when an alternating signal voltage is applied to the control circuit in place of a direct-voltage signal are more fully appreciated.

Special effects observed in the a.c. control of single-element transductors are discussed, and consideration is given to 2-element

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Milnes and Mr. Law are at the Royal Aircraft Establishment.

circuits for full-wave output and to various push-pull arrangements.

Other possible amplifier arrangements responding to signals of the same frequency as the supply may be mentioned. A phase-sensitive rectifier may be used to produce a rectified output of polarity dependent on the signal phase, and this output may then be applied to conventional d.c.-controlled transductor arrangements. A second method of handling an alternating-voltage signal is to apply one half-cycle to one transductor and the other half-cycle to a second transductor by means of routing rectifiers and then to combine the two transductor outputs so that the required form of resultant output (a.c. or d.c.) is obtained. This technique has been applied by Maine to reset transductors of the Ramey type to produce a.c. or d.c. push-pull outputs with alternating-voltage control signals.[3]

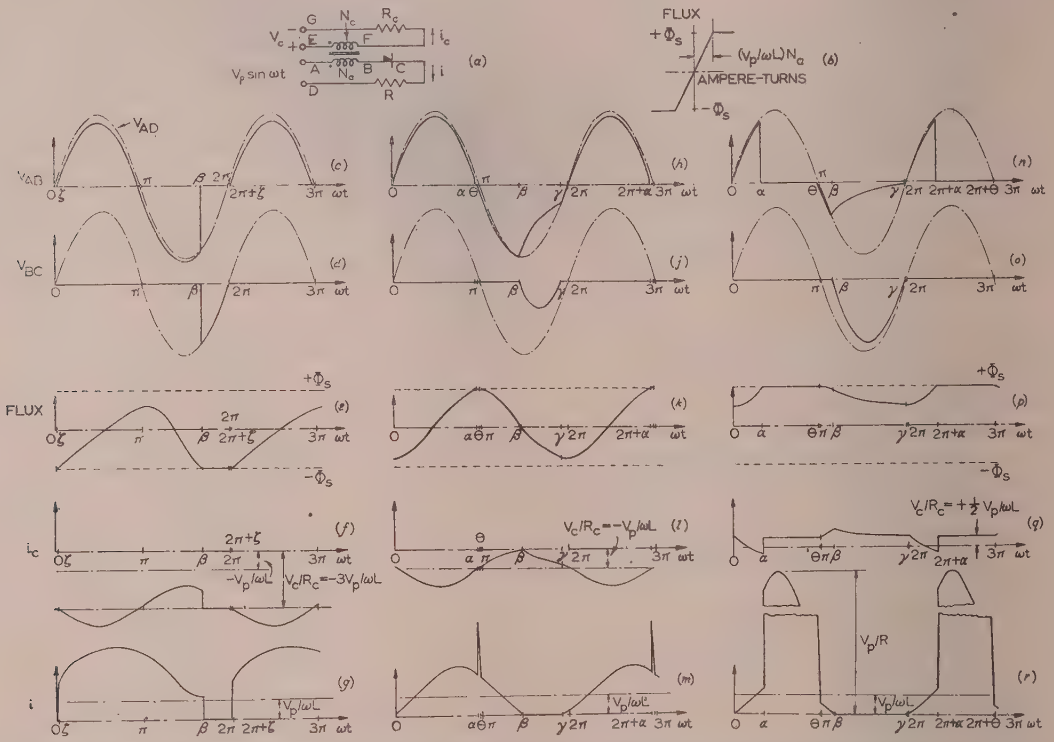


Fig. 2.—Theoretical waveforms for single auto-self-excited element with direct-voltage signal source of finite resistance.

The waveforms are drawn for the full supply-voltage rating of the element. The control turns are equal to the main-winding turns, the control-circuit resistance is equal to the control-winding reactance and the load resistance is taken as one-thirtieth of the main-winding reactance. Control-voltage conditions are as follows:

(c)–(g) $V_c/R_c = -3V_p/\omega L$,
(h)–(m) $V_c/R_c = -V_p/\omega L$,
(n)–(r) $V_c/R_c = +\frac{1}{2}V_p/\omega L$.

(a) Single-element circuit.
(b) Assumed core-magnetization characteristic.
(c), (h), (n) Supply and voltage across main winding.
(d), (j), (o) Voltage across circuit rectifier.
(e), (k), (p) Core flux.
(f), (l), (q) Current in control circuit.
(g), (m), (r) Current in load.

(2) SINGLE-ELEMENT BEHAVIOUR WITH DIRECT-VOLTAGE CONTROL

The theoretical treatment which follows for the single-element circuit is idealized by the assumption of a core B/H characteristic that is a straight line, i.e. of constant permeability, up to the level of complete flux saturation. No attempt is made to take account of hysteresis in the assumed saturation characteristic since this would add appreciably to the complexity of the analysis, and the linear B/H characteristic assumed is quite sufficient to yield instructive characteristics and waveforms not far removed from those observed in practice.

Further simplifications made in the treatment are that the winding resistances are negligible and that the auto-self-excitation rectifier has negligible forward resistance and infinite reverse resistance.

(2.1) Circuit Action

For the single-element circuit with direct-voltage control there are two main modes of action to be considered. One relates to the high-slope region of the characteristic used for amplification and the other to the low-slope region that occurs at large negative control signals. These modes of action are illustrated by the theoretical waveforms in Fig. 2, and the particular circuit conditions that these represent are stated in the captions.

For a positive control current equal to half the saturation-knee current of the assumed core magnetization characteristic [Fig. 2(b)], the corresponding waveforms are given in Figs. 2(n) to 2(r). Consider, the cycle of events occurring between α and $2\pi + \alpha$, in sequence, commencing with the period α to θ during which the core is saturated in the positive sense. Since the core is saturated the whole of the supply voltage appears across the load, and the load current is therefore of sinusoidal form $(V_p \sin \omega t)/R$ between the limits α and θ. (In the waveform in Fig. 2(r) the amplitude of the load current in this region has had to be shown in compressed form for ease of representation.) During this period no voltage is induced into the control circuit from the main circuit because the core is saturated, and therefore the control-circuit current is determined directly by the signal voltage and the circuit resistance and is constant as shown in Fig. 2(*q*).

At the angular position θ, shortly before π, the main circuit excitation has fallen to a value which makes the total excitation on the core equal to that of the knee of the saturation characteristic, and the core therefore desaturates. Between θ and β the core sustains negative voltage and the main circuit current decreases as shown in Fig. 2(*r*). In the control circuit the induced voltage causes the control current to increase, as in Fig. 2(*q*), as may be seen by observing the circuit-winding senses and current directions marked in the circuit diagram of Fig. 2(*a*).

At the angular position β the current in the main circuit has become zero and the voltage across the main winding is just equal to the supply voltage as shown in Fig. 2(*n*). Since the current is zero it is then possible for the circuit rectifier to begin to sustain reverse voltage, which it does between β and γ, as in Fig. 2(*o*). During this period the main circuit current is zero but the voltage across the core main winding changes exponentially. In the control circuit this induces an exponential decay of current, which, because of the change of core excitation, corresponds to an exponential decrease of core flux, and on differentiation this accounts for the exponential decrease of element voltage.

The interval of rectifier reverse voltage continues until the angular instant γ at which the supply voltage has decreased to the value of the element voltage in its process of exponential decay. Then follows a period γ to $2\pi + \alpha$ during which magnetizing current again flows in the main circuit and the core sustains a positive voltage and the core flux increases. The control-circuit current during this period tends to dip as shown in Fig. 2(*q*) because the element voltage induced on the circuit opposes the signal voltage. The core flux increases until saturation at $2\pi + \alpha$, which completes the cycle of events.

Waveforms Figs. 2(*h*) to 2(*m*) show the same mode of action at a very retarded triggering angle α, corresponding to a low transductor output—the mean control current for the condition illustrated corresponds to the negative saturation knee condition (i.e. $V_c/R_c = -V_p/\omega L$). From the waveforms of Figs. 2(*l*) and 2(*m*) it is seen that the control and main circuit currents may exceed the saturation knee values, but the resultant combined excitation at all times corresponds to the flux and must not of course exceed the knee excitation if the core is unsaturated.

The other principal mode of action is that on the low-slope region of the characteristic for conditions of large negative signal. This is illustrated by the waveforms Figs. 2(*c*) to 2(*g*) in which it is seen that the core now saturates in the negative flux direction during a period β to $2\pi + \xi$. The supply voltage is then sustained as rectifier reverse voltage [see Fig. 2(*d*)]. The control-current and main-current waveforms are as shown in Figs. 2(*f*) and 2(*g*) with a resultant value corresponding to the core flux variation shown in Fig. 2(*e*). It will be seen that the core reaches negative saturation at β, but the main circuit current is not zero at that instant and therefore a sharp cut-off of the main current must occur at β as the rectifier takes up reverse voltage. Since the element voltage is zero when saturated, the control current at β jumps to the value given by V_c/R_c.

Between this mode of action and the one discussed previously [Figs. 2(*h*) to 2(*m*)] there is a transitional mode of action in which the main circuit current reaches zero before the core flux saturates. This has been examined, but it is not of sufficient importance to justify discussion in detail.

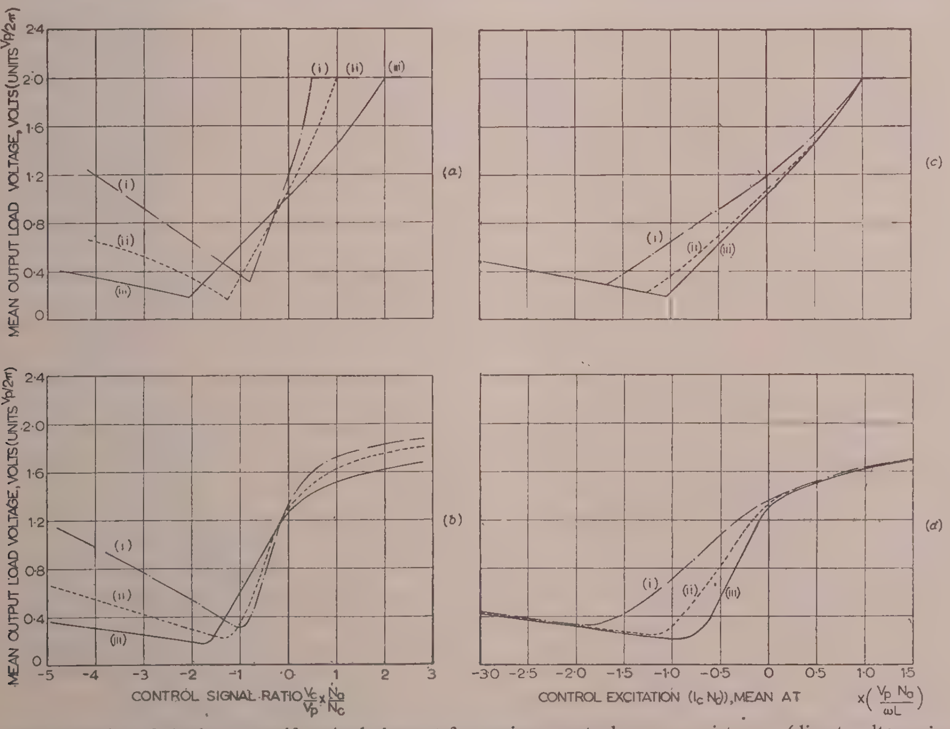


Fig. 3.—Characteristics of single auto-self-excited element for various control-source resistances (direct-voltage signal).

(*a*), (*c*) Theoretical characteristics.
(*b*), (*d*) Practical curves for a core of Mumetal.

(i) Curves for $R_c = \frac{1}{2}\omega L_c$.
(ii) Curves for $R_c = \omega L_c$.
(iii) Curves for $R_c = 2\omega L$.

All curves are for a load resistance of one-thirtieth of the main-winding unsaturated reactance.

(2.2) Theoretical and Practical Characteristics

The analytical treatment follows directly from the modes of action described. For the action represented by the waveforms of Figs. 2(*n*) to 2(*r*) the equations for the various circuit conditions are given in Appendix 10.1. For large negative signals the mode of action is represented by the waveforms of Figs. 2(*c*) to 2(*g*), and the analytical expressions are summarized in Appendix 10.2. From the analytical expressions given, theoretical characteristics of mean load-voltage/signal-voltage or current have been calculated and are presented in Figs. 3(*a*) and 3(*c*). Decrease of control-circuit resistance is seen to increase the sensitivity with respect to voltage for both limbs of the characteristic [see Fig. 3(*a*)], but to require more signal current for a given output change [Fig. 3(*c*)].

To confirm these theoretical characteristics, experiments were made with a single-element transductor having a core of Mumetal overlapping-E laminations, at 400 c/s supply frequency. With the circuit proportions of resistance to reactance chosen to be about the same as those assumed in the calculations, characteristics as in Figs. 3(*b*) and 3(*d*) were obtained. The agreement between the theoretical and experimental characteristics in form is reasonable, except in the region of full output, where the practical curves have curvatures corresponding to the gradual saturation knee of the Mumetal.

As further confirmation of the treatment, the circuit waveforms were examined and found to be similar in general form to those predicted theoretically.

(3) ALTERNATING-VOLTAGE CONTROL OF SINGLE-ELEMENT TRANSDUCTOR

(3.1) Introduction

The treatment of the single-element circuit for direct-voltage control lays a foundation from which it is possible to proceed to the discussion of alternating-voltage control. Here it is necessary to take account of the phase of the signal with respect to that of the supply. In Fig. 4(*a*) the in-phase condition is defined as that for which the polarities on A and E (the starts of the two windings, as indicated by the dot symbol) are the same.

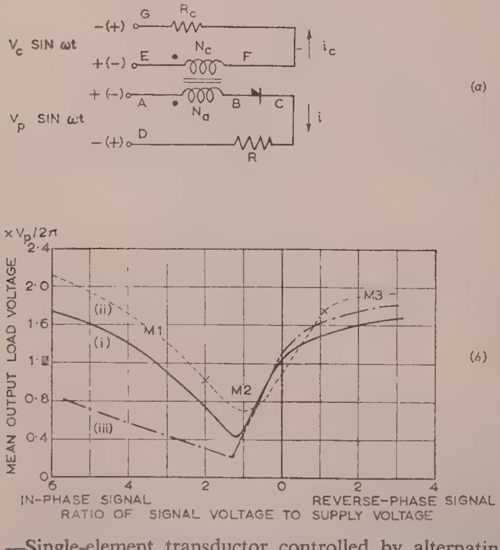


Fig. 4.—Single-element transductor controlled by alternating signal voltage of the same frequency as the main circuit supply.

(*a*) Circuit diagram. The instantaneous polarities shown for the supply and control voltages are defined as the in-phase condition.
(*b*) Typical characteristics for equal main and control windings.

(i) Practical curve for Mumetal-cored element.
(ii) Theoretical curve for alternating-voltage control.
(iii) Practical curve for direct-voltage control.

A brief indication of the type of characteristic obtained with alternating-voltage control is perhaps useful before a detailed discussion of the circuit behaviour is undertaken. A typical experimental characteristic with in-phase and reverse-phase signals is therefore given as curve (i) in Fig. 4(*b*). A particular feature of this curve is the rise in output that occurs for in-phase signals of large amplitude. With direct-voltage control [Fig. 4(*b*) curve (iii)], the analogous condition of large negative-signal input causes much less rise in output. It will be appreciated that in the a.c.-controlled circuit the appreciable response to large in-phase signals is inconvenient in the use of the characteristic for phase-sensitive amplification.

Curve (ii) of Fig. 4(*b*) represents a theoretical characteristic for in-phase and reverse-phase control based on the straight-line idealization of the core saturation characteristic, as in the previous analytical treatment. The theoretical characteristic lies above the practical curve [Fig. 4(*b*) curve (i)], but this arises mainly from the voltage loss across the winding and the rectifier forward-resistances in the main circuit of the experimental transductor.

(3.2) Control Action for In-phase and Reverse-phase Signals

For a single-element transductor circuit with in-phase and reverse-phase control signals there are three modes of action to be considered. In Fig. 4(*b*) curve (ii) the regions of the characteristic corresponding to these modes of action are indicated by the symbols M1, M2 and M3.

The action of the circuit is illustrated in Fig. 5, where the waveforms (*c*)–(*g*) are for an in-phase signal of three times the supply-voltage amplitude (M1 operation), and the waveforms (*h*)–(*m*) are for an in-phase signal equal in amplitude to that of the supply voltage (M2 operation). The remaining waveforms, (*n*)–(*r*), are for a reverse-phase signal of supply-voltage amplitude which is still in M2 operation near the point where transition from M2 to M3 operation occurs.

Consider initially waveforms (*c*)–(*g*) commencing from the position α at which the core is just saturated. This saturated condition continues practically up to π, and the load current [Fig. 5(*g*)] represents the supply voltage appearing across the load resistance. Since the element voltage is zero during the period, the control current [Fig. 5(*f*)] corresponds to the full signal voltage applied across the control resistance. Almost at π the core desaturates, and the voltage across the main winding during the ensuing period π to β is just slightly greater than the negative half-cycle of the supply voltage. This difference which exists because of a small voltage drop across the load produced by a small positive current flowing in the main circuit, as in Fig. 5(*g*), is explained below.

During this period π to β the voltage across the control-circuit resistance is the difference between the signal and element voltages, and a corresponding current flows in the circuit. The total excitation on the element must, however, correspond to the flux changes in the core [Fig. 5(*e*)], and since the control-circuit current is too negative for that purpose, the required sum is made up by a positive magnetizing current flowing in the main circuit as shown in Fig. 5(*g*). This current becomes zero at β and the circuit rectifier then begins to sustain reverse voltage. The control of the transductor is then taken over by the signal circuit, the equation to be satisfied being as follows:

$$V_c \sin \omega t = L_c \frac{di_c}{dt} + i_c R_c \quad . \quad . \quad . \quad . \quad (14)$$

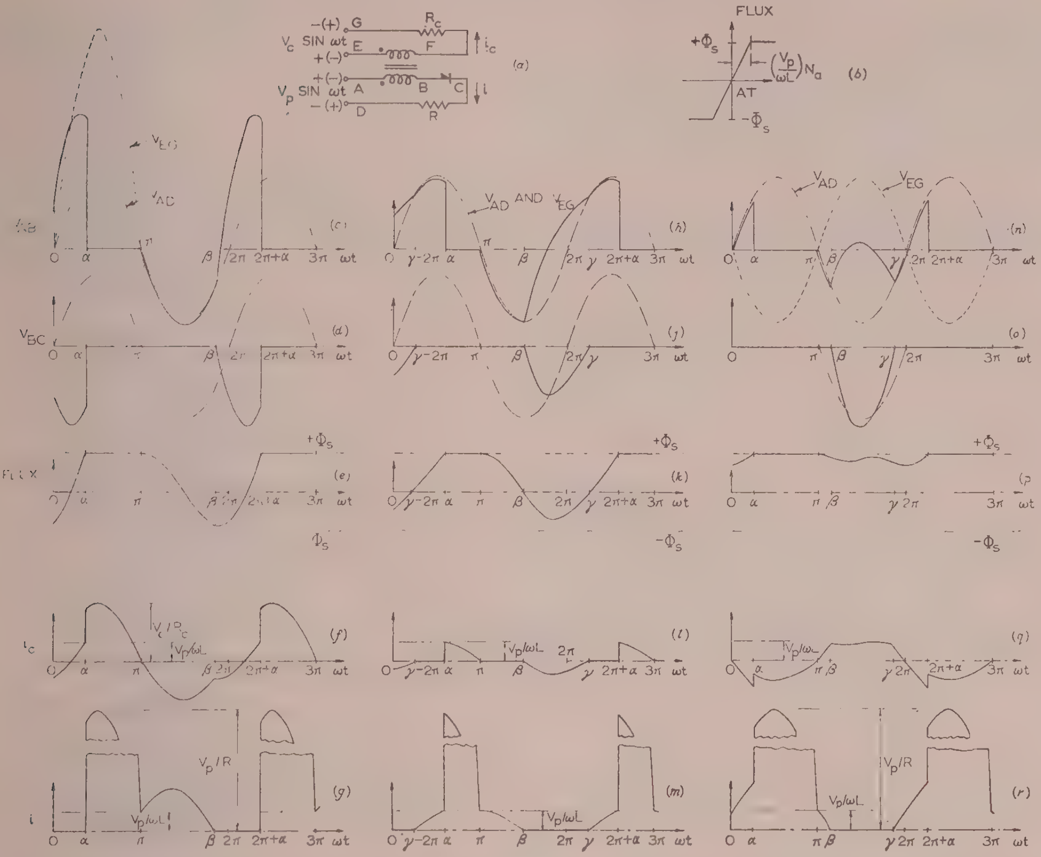


Fig. 5.—Theoretical waveforms for single auto-self-excited element with alternating-voltage signal source of finite resistance.

Drawn for the conditions $\omega L_c/R_c = 1$, $\omega L/R = 30$, $N_a/N_c = 1$ with supply voltage equal to the full rated value for the element.

Control-voltage conditions are as follows:
(c)–(g) $V_c/V_p = 3$,
(h)–(m) $V_c/V_p = 1$,
(n)–(r) $V_c/V_p = 1$ with 180° phase difference between control and supply voltages.

(a) Circuit diagram.
(b) Assumed flux characteristic.
(c), (h), (n) Supply and control voltages and voltage across main winding.
(d), (j), (o) Voltage across rectifier.
(e), (k), (p) Core flux.
(f), (l), (q) Current in control circuit.
(g), (m), (r) Current in load.

Since the control signal is large, appreciable voltage is developed across the element during the period, and the core flux increases until the positive saturation level is reached at the instant $2\pi + \alpha$. The distribution of voltage and current which results in the circuit during the period β to $2\pi + \alpha$ is shown in the waveforms Figs. 5(c) to 5(g).

Consider now the mode of action for an in-phase signal equal to the supply voltage, this condition being represented by the waveforms Figs. 5(h) to 5(m). Here the action is similar to that described above except that during the period following β the signal voltage loses sole control of the transductor at an angular position γ. The rectifier then begins to conduct again, and the main circuit and signal circuits share the control of the transductor until the core flux is carried into positive saturation at $2\pi + \alpha$.

For a reverse-phase signal equal to the supply voltage, the waveforms are as in Figs. 5(n)-5(r). The waveshapes differ rather from those in Figs. 5(h)–5(m), because the rectifier conducts at an angle γ which is earlier than 2π, but the mode of action is basically the same. In Fig. 5(n) it will be seen that the element voltage is positive for a portion of the period β to γ, and this produces a corresponding rise in the flux waveform, Fig. 5(p). For larger signal voltages this rise of flux becomes sufficient to produce a further period of core saturation. This third mode of circuit action has been fully examined, but it will not be illustrated and discussed in detail here because, as will be seen from Fig. 4(b) curve (ii), it represents a region of the characteristic that is not as important as those covered by the other two modes of behaviour.

For the modes of action that have been described, the circuit behaviour may be treated analytically and expressions derived for the load voltage and the angular parameters in terms of the signal voltage. In Appendix 10.3 the derivation of these expressions is given for the mode of action represented by the waveforms of Figs. 5(h) to 5(m).

In the use of these expressions for the calculation of theoretical output characteristics, the first stage is to determine β from eqn. (20) for a given signal voltage. The next step is to solve eqns. (22) and (23) in order to determine the corresponding value of γ, and the mean output voltage is then calculable from eqns. (19) and (21).

The treatment of the mode of transductor action represented by the waveforms of Figs. 5(c) to 5(g) was made more rigorous by the inclusion in the analysis of terms for voltages appearing across the load resistance as a result of magnetizing components of main circuit current. This refinement was advisable because these magnetizing components are rather larger for this mode of action than for the others.

(3.3) Action with Phase-Shifted Signals

General circuit equations for the action when the signal voltage is displaced relative to the supply voltage by a phase shift, say ψ, may be written down and developed in the same way as for the zero-phase-shift condition. The equations connecting the angular parameters are, of course, rather more cumbersome because of the terms in ψ and will not be given here.

One condition which has been fully examined, and solved numerically, is that for a $\pi/4$ phase lag (or on phase reversal a $3\pi/4$ lead) with the control resistance equal to the control winding reactance. The theoretical waveforms calculated for this particular condition are given in Fig. 6, and should be compared with those given in Fig. 5 for zero phase shift.

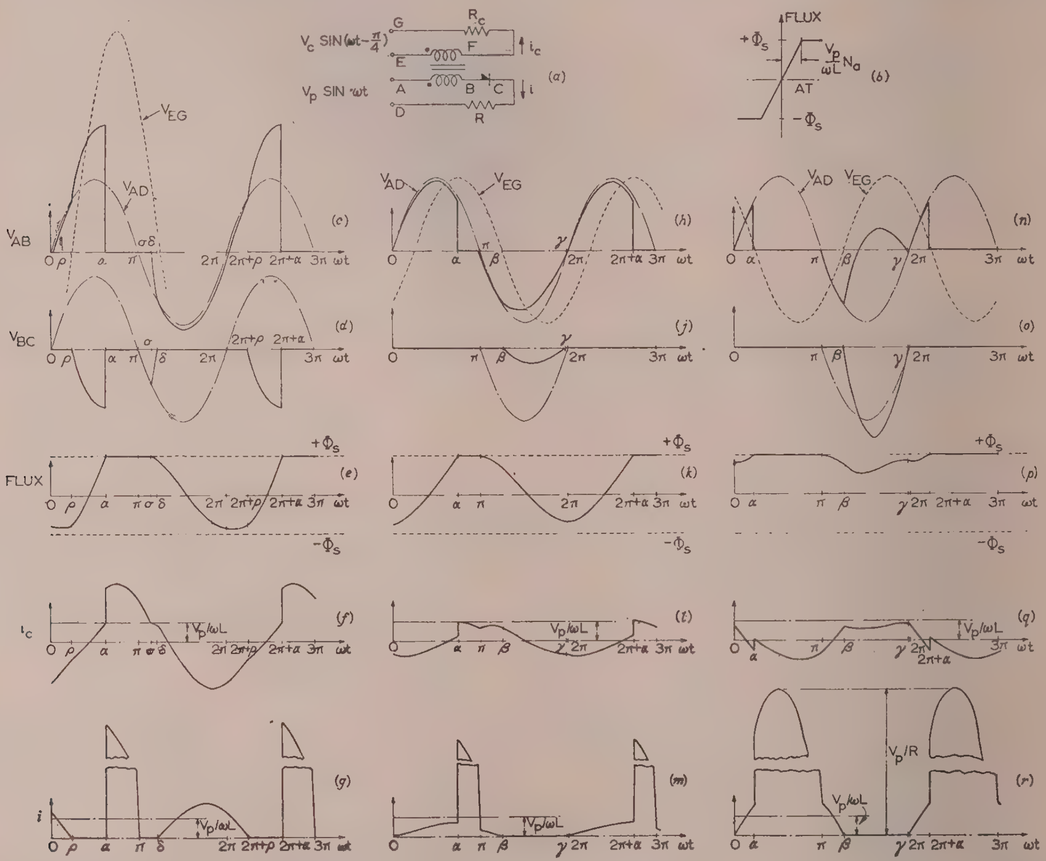


Fig. 6.—Theoretical waveforms for single auto-self-excited element with alternating control voltage lagging $\pi/4$ behind the supply voltage or leading by $3\pi/4$ on phase reversal.

Drawn for the conditions $\omega L_c/R_c = 1$, $\omega L/R = 30$, $N_a/N_c = 1$.
Control-voltage conditions are as follows:
(c)–(g), $V_c/V_p = 3$,
(h)–(m), $V_c/V_p = 1$,
(n)–(r), $V_c/V_p = 1$ with reversal of phase.

(a) Circuit diagram.
(b) Assumed flux characteristic.
(c), (h), (n) Supply and control voltages and voltage across main winding.
(d), (j), (o) Voltage across rectifier.
(e), (k), (p) Core flux.
(f), (l), (q) Current in control circuit.
(g), (m), (r) Current in load.

For the signal condition $V_c/V_p = 1$ it will be seen from Figs. 5(m) and 6(m) that the principal effect of the phase shift is to increase the range of movement of the triggering angle α and so to improve the swing in output voltage.

Comparison of the waveforms (c)–(g) of Figs. 6 and 5 shows that the phase-lag of $\pi/4$ has the effect of reducing the output obtained for the signal amplitude $V_c = 3V_p$. Since this is in the nature of a swamp signal, the reduction in output response is a desirable effect. It will be noticed that a second period of rectifier cut-off, between π and δ, occurs in the circuit action, with phase lag present. Features of this kind, involving variation of the circuit action, make the presentation of a full treatment for a perfectly general phase shift ψ rather involved and hardly worth while.

(3.4) Theoretical and Practical Characteristics

The action of the single-element transductor with alternating-voltage control has now been discussed in detail for the zero-phase-shift and $\pi/4$ phase-lag conditions. The corresponding characteristics are curves (i) and (ii) in Fig. 7(a), and an addition to these is curve (iii), which is the theoretical characteristic for the $\pi/4$ phase-lead condition. The features of these characteristics which should be noted are the limited range of load-voltage change obtainable in curves (i) and (iii) and the V-shapec form of all three curves.

Practical characteristics for a Mumetal-cored transductor are given in Fig. 7(b), and it will be seen that the curves for in-phase and $\pm\pi/4$ phase-shift conditions have forms that agree generally with the shapes predicted by the theoretical treatment. Curve (iv) in Fig. 7(b) is for the $-\pi/2$ phase shift condition, and the $+\pi/2$ phase-shift curve is of course symmetrical.

The remaining characteristic [Fig. 7(b) curve (v)] is for direct-

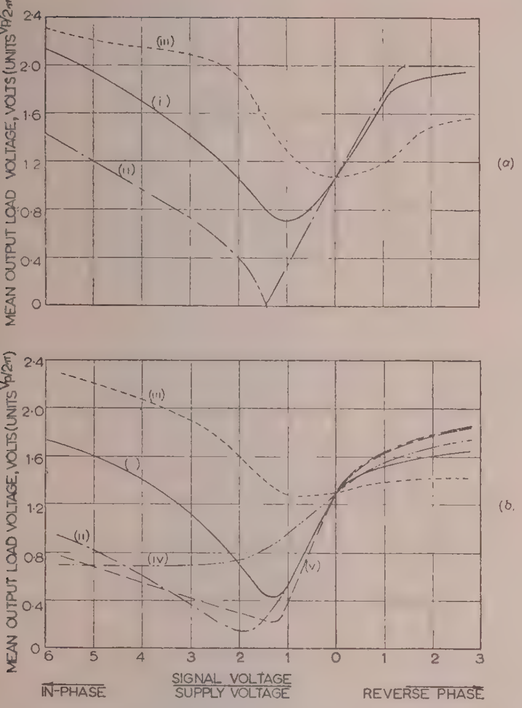


Fig. 7.—Characteristics of single auto-self-excited element (resistive load) with control-voltage phase shift.

(*a*) Theoretical characteristic for $\omega L_c/R_c = 1$, $\omega L/R = 30$, $N_a/N_c = 1$.
(*b*) Practical characteristic for $\omega L_c/R_c = 1$, $\omega L/R = 30$, $N_a/N_c = 1$.

(i) No phase shift.
(ii) $\psi = -\pi/4$.
(iii) $\psi = +\pi/4$.
(iv) $\psi = -\pi/2$.
(v) Characteristic with direct voltage applied to the control circuit.

voltage control of the element and is given for purposes of comparison. For this curve, the abscissa scale is the ratio of the direct-voltage signal to the peak-value of the supply voltage, whereas for alternating-voltage control the scale is the ratio of the peak or r.m.s. values.

The waveforms of the practical transductor were found to be in good agreement with those predicted theoretically for the various alternating-voltage-control signal conditions.

An inductance connected in series with the control winding was found to produce an alteration in the shape of the output characteristic rather similar, though not completely so, to that produced by phase lag of the signal voltage. Series capacitance in the control circuit (with $1/\omega C$ about equal to ωL_c) reduces the standing output at zero signal and increases the sensitivity if the signal voltage lags on the supply by $\pi/2$.

(4) DISCUSSION OF ALTERNATING-VOLTAGE CONTROL

(4.1) Influence of Control-Circuit Parameters

For an understanding of the characteristics of a single-element transductor, the effects of changing either the number of control-winding turns or of the control-circuit resistance need discussion.

Clearly the signal voltage must be equal to the sum of the voltage appearing across the control winding of the element and the voltage appearing across the control circuit resistance as a result of the current flowing in the circuit. For simplicity consider a condition of zero output such as occurs in Fig. 7(*a*) curve (ii). The transductor main winding then sustains the whole of the supply voltage, and by normal transformer action the voltage across the control winding is $(N_c/N_a)V_p \sin \omega t$. The main circuit current is completely zero for this particular circuit condition, and the core excitation corresponding to the voltage across the element is supplied entirely by the control circuit. From the assumed magnetization characteristic of the core [Fig. 6(*b*)], the necessary control current is therefore

$$i_c = \frac{N_a}{N_c}\frac{V_p}{\omega L}\sin(\omega t - \pi/2) \quad . \quad . \quad . \quad (24a)$$

or

$$= \frac{N_c}{N_a}\frac{V_p}{\omega L_c}\sin(\omega t - \pi/2) \quad . \quad . \quad . \quad (24b)$$

and when this is multiplied by R_c it gives the voltage appearing across the control resistance. The required signal voltage is the vector sum of the two voltages in the control circuit and is therefore given by

$$v_c = \frac{N_c}{N_a}V_p\left[1 + \left(\frac{R_c}{\omega L_c}\right)^2\right]^{1/2}\sin(\omega t - \psi) \quad . \quad . \quad (25)$$

where $\psi = \arctan(R_c/\omega L_c)$

Thus, for example, if $R_c = \omega L_c$ the theoretical signal-voltage ratio for minimum output is $\sqrt{(2)}N_c/N_a$ and the required signal lag is $\pi/4$, as in Fig. 7(*a*) curve (ii).

Consider now the experimental characteristics shown in Fig. 8, which are for a Mumetal-cored transductor with no signal phase-shift. The discussion given above for the zero output condition is, of course, not fully applicable to these characteristics, but nevertheless it is a useful guide in interpreting the circuit behaviour. Comparison of curves (i) and (ii) of Fig. 8 shows that doubling the number of control-winding turns has the expected effect of almost doubling the signal voltage required for minimum output. However, comparison of curves (i) and (iii) shows that the signal voltage for minimum output is not greatly changed by a reduced control winding. The explanation is that the control current required with the reduced winding is increased, the voltage across the control resistance increasing accordingly, and with certain circuit proportions this may tend to mask the reduction in the voltage induced in the control circuit by transformer action from the main circuit.

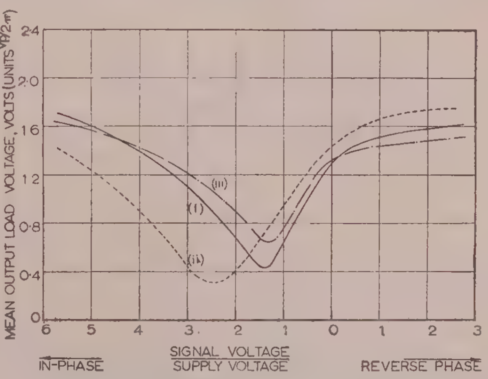


Fig. 8.—Experimental characteristics to show the effect of change of control turns with the control resistance constant.

(i) $N_c = N_a$.
(ii) $N_c = 2N_a$.
(iii) $N_c = \frac{1}{2}N_a$.

From the explanation so far given it may be concluded that the control-circuit resistance causes a loss of sensitivity by its presence. The practical characteristics given in Fig. 9(*a*) confirm this, but from Fig. 9(*b*) it will be appreciated that the resistance R_c does serve the useful function of limiting the current drawn

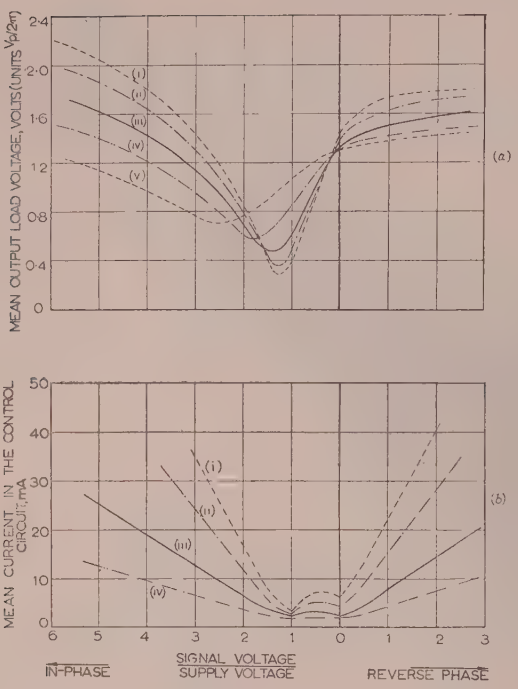


Fig. 9.—Experimental characteristics for single-element transductor with change of control-circuit resistance.

(i) $R_c =$ 500 ohms corresponding to $\omega L_c/R_c = 3$.
(ii) $R_c =$ 750 ohms corresponding to $\omega L_c/R_c = 2$.
(iii) $R_c =$ 1 500 ohms corresponding to $\omega L_c/R_c = 1$.
(iv) $R_c =$ 3 000 ohms corresponding to $\omega L_c/R_c = 0{\cdot}5$.
(v) $R_c =$ 6 000 ohms corresponding to $\omega L_c/R_c = 0{\cdot}25$.

Transductor details. Supply voltage 10 volts (r.m.s.) 400 c/s.
$N_a = N_c = 200$.
Mumetal core. Load 47 ohms.

from the source or resulting from the voltage induced in the control circuit from the main circuit.

In design work, the minimum value of R_c is determined by the maximum permissible current loading for the control-signal source, which may be a synchro or some other form of alternating-voltage pick-up unit sensing an angular or linear movement. In the transductor design the problem is then to achieve a high ratio of main- to control-winding turns, N_a/N_c, since this gives voltage amplification, while at the same time maintaining N_c large since this reduces the control-circuit magnetizing current. If this current is too large it causes a voltage drop across R_c which has the effect of increasing the signal voltage required. From eqn. (25) it is seen that preferably the ratio $R_c/\omega L_c$ should be unity or less, if the signal voltage is to be suitably limited.

(4.2) Core Material and Design Flux Density

Curves (i)–(iii) of Fig. 10(*a*) show the effect of increasing the number of main-winding turns of a single-element transductor and increasing the supply voltage in proportion. These related changes of supply voltage and turns ratio result in no change in the voltage induced in the control circuit, and therefore, as expected, there is no alteration in the signal voltage required for minimum output in curves (i)–(iii). Increase of the turns ratio and supply voltage does, of course, increase the circuit gain since the output power, or voltage, controlled for a given signal-voltage change is increased.

The principles governing the matching of the load resistance to the transductor main-circuit internal resistance are the same as for conventional d.c.-controlled circuits. Briefly, these are that with the load equal to the internal resistance high power gain is obtained, but the heat-dissipation rating of the transductor then represents the limiting power that can be achieved in the load since the circuit efficiency is only 50%. If higher efficiency is required, the circuit may be designed with a greater ratio of load to internal circuit resistance.

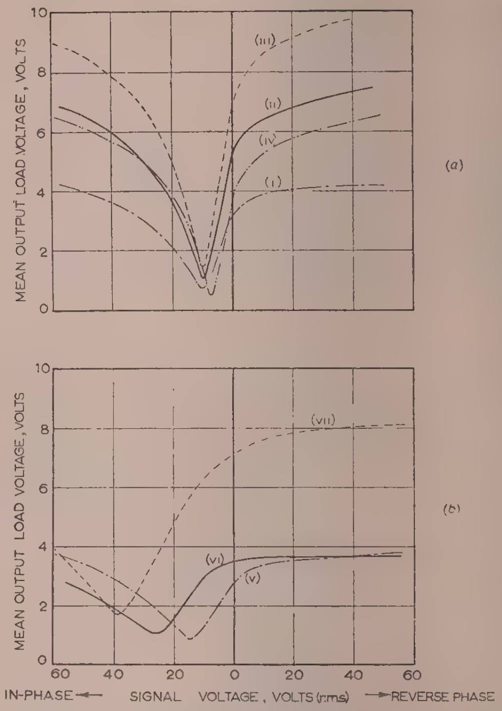


Fig. 10.—Effect on single-element characteristic of changing supply voltage and core material.

(*a*) Increase in N_a and supply voltage with control circuit unchanged.
$N_c = 141$; $R_c = 750$ ohm; $R = 47$ ohms; Mumetal core.
(i) $N_a = 200$. Supply voltage = 10 volts (r.m.s.).
(ii) $N_a = 400$. Supply voltage = 20 volts (r.m.s.).
(iii) $N_a = 600$. Supply voltage = 30 volts (r.m.s.).
(iv) $N_a = 600$. Supply voltage = 20 volts (r.m.s.).

(*b*) Effect of changing core material.
$N_a = 200$; $N_c = 200$; $R_c = 1\,500$ ohms; $R = 47$ ohms.
(v) Mumetal core, 10 volts (r.m.s.) supply.
(vi) H.C.R. core, 10 volts (r.m.s.) supply, stack depth reduced to $\frac{5}{11}$ of (v).
(vii) H.C.R. core, 22 volts (r.m.s.) supply, stack depth as in (v).

Curves (iii) and (iv) of Fig. 10(*a*) are for the same main winding at supply voltages corresponding to the normal design flux density for Mumetal transductors of 5 000 gauss (peak), and to a reduced flux density of 3 300 gauss (peak). At the lower supply voltage, the signal and output voltages are both reduced and the net effect is a slight loss of sensitivity. This behaviour is rather similar to that for d.c.-controlled transductors, and the design flux densities normally used for these are equally suitable for a.c.-controlled circuits.

The rectangular-loop 50% nickel–iron core materials, of which H.C.R. alloy is one, have considerable use in transductor technique since they may be operated at design flux densities of 11 000 gauss (peak), or even more, and so require less core cross-sectional area than Mumetal at the same controlled power output. Curves (v) and (vi) of Fig. 10(*b*) are for Mumetal and H.C.R. transductors at the same supply voltage with the lamination stacks in the ratio 11/5 so that both materials are operating at their normal design flux densities. The two curves are very similar in shape and sensitivity, but the H.C.R. characteristic is displaced to the left of the other curve, presumably because of the appreciably wider hysteresis, on total-loss loop, of the material. Curve (vii) is for an H.C.R.-cored element having the same stack depth as the Mumetal one, and with the circuit supply voltage appropriately increased. The sensitivity is seen to be very nearly the same as for the other two curves in Fig. 10(*b*), but the output voltage and power controlled are, of course, considerably greater. For low power-output designs Mumetal-type core materials are perhaps more suitable than the rectangular-loop materials, which in the form of very small cores tend to be more sensitive to strain effects on the magnetic properties.

(4.3) Behaviour with Bias Excitation

For the single-element transductor with a third winding used for bias-excitation purposes, characteristics as shown in Figs. 11(*a*) and 11(*b*) are obtained. Negative d.c. bias excitation introduces a change in shape of the characteristic [Fig. 11(*a*)] by insertion of a region of low output between the two sides of the curve. Positive d.c. bias is not of interest, because the minimum output of the transductor rises and there is virtually no signal action on the output. With a.c. bias [Fig. 11(*b*)] the effect is to displace the characteristic along the signal-voltage axis without change of shape. For these measurements the bias circuit was maintained at high impedance relative to the control circuit, i.e. $\omega L_b/R_b$ was considerably less than $\omega L_c/R_c$, so that the bias current waveforms would not be seriously affected by the voltage across the element.

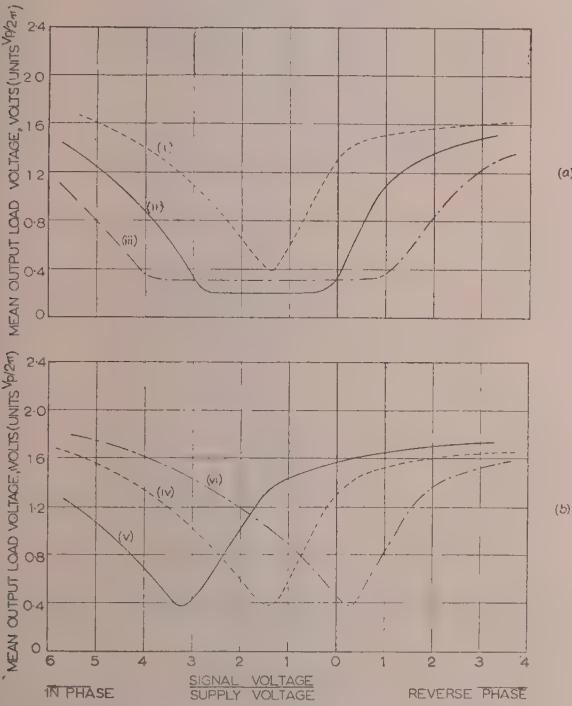


Fig. 11.—Effect of d.c. and a.c. bias on the characteristics of an a.c. controlled single-element transductor.

Test conditions $N_a = N_b = N_c$; $\omega L/R = 30$; $\omega L_c/R_c = 1$; $\omega L_b/R_b = 0{\cdot}15$; Mumetal core. Supply voltage = 10 volts (r.m.s.) 400 c/s.

(*a*) D.C. bias obtained from a direct-voltage supply.
(i) Zero bias.
(ii) −10 mA bias.
(iii) −20 mA bias.

(*b*) A.C. bias without rectification.
(iv) Zero bias.
(v) 10 mA (r.m.s.) bias current for in-phase bias voltage.
(vi) 10 mA (r.m.s.) bias current for reverse phase bias voltage.

Rectified half-wave bias was next investigated, and some typical characteristics together with the circuit arrangement are given in Fig. 12(*a*). With negative half-wave bias applied in the load-conducting half-cycle [Fig. 12(*a*) curves (ii) and (iii)], the effect is seen to be a shift of the left-hand limb of the characteristic, whereas with the bias applied in the core-reset half-cycle [Fig. 12(*a*) curve (iv)], the right-hand limb is displaced.

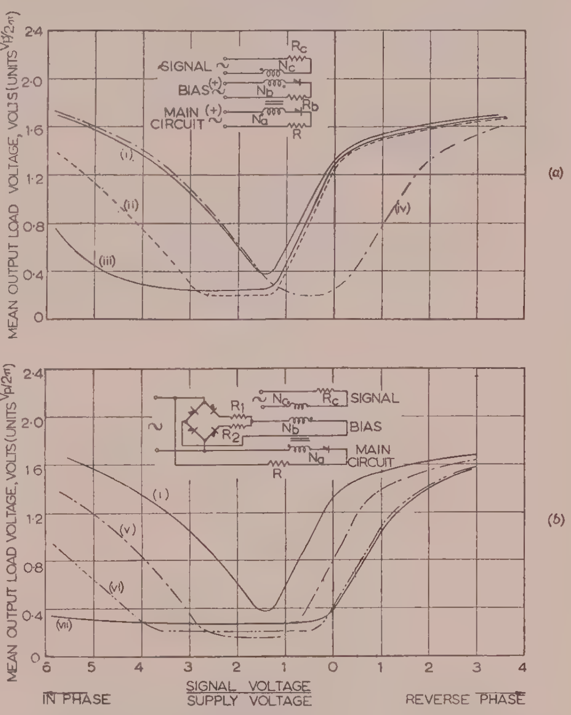


Fig. 12.—Characteristics for other forms of bias excitation.

Test conditions as for Fig. 11.

(*a*) Half-wave rectified bias.
(i) Zero bias.
(ii) −5 mA (mean) bias in the load-conducting half-cycle.
(iii) −10 mA (mean) bias in the load-conducting half-cycle.
(iv) −5 mA (mean) bias in the resetting half-cycle.

(*b*) Full-wave rectified bias and also divided-bridge bias.
(i) Zero bias.
(v) −5 mA (mean) with bridge complete and single resistance of 3·85 kilohms.
(vi) −10 mA (mean) with bridge complete and single resistance of 1·88 kilohms.
(vii) −10 mA (mean) through $R_2 = 1{\cdot}35$ kilohms.
−5 mA (mean) through $R_1 = 1{\cdot}79$ kilohms (divided-bridge bias).

With full-wave rectified bias [Fig. 12(*b*) curves (v) and (vi)] both limbs of the characteristic are displaced and the whole effect is very similar to that obtained with smooth d.c. bias [Fig. 11(*a*)]. The rise in output for large in-phase signals may be more completely displaced by the application of greater bias current in the load-conducting half-cycle than in the reset half-cycle. This is illustrated by the circuit inset in Fig. 12(*b*) and by curve (vii). For convenience of reference in later Sections, this circuit variation is termed divided-bridge bias.

(5) FULL-WAVE CIRCUIT ARRANGEMENTS WITH A.C. CONTROL

Single-element transductor behaviour having been examined, the next stage is the discussion of full-wave circuit arrangements, for which two elements are required.

The main windings of the 2-element circuits may be connected in any of the ways usual for d.c.-controlled transductors, i.e. auto-self-excited parallel, bridge or centre-tap arrangements.[1] The control windings may be connected either in parallel or in series. If series connected, the circuit action requires that the windings shall be in the sense that results in summation of the induced voltages from the main circuit, instead of in the cancellation sense necessary in d.c.-controlled transductors. Thus a full-wave transductor which responds to signals of supply frequency does not provide usable characteristics for d.c. control, and vice versa, unless the control circuit is rearranged for the changed signal input.

(5.1) Characteristics for Parallel and Series-Connected Control Windings

The characteristics obtained with the control windings of a 2-element circuit connected in parallel and in series are rather different. In Figs. 13(*a*) and 13(*b*) the two control-circuit arrangements are shown for a transductor with the main windings in auto-self-excited parallel connection, and typical characteristics are given in Fig. 13(*c*) together with a curve for single-element half-wave operation.

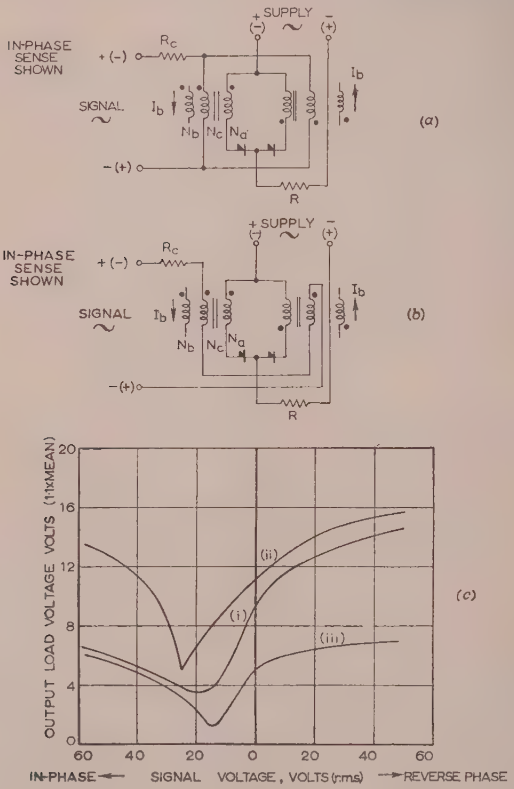


Fig. 13.—Auto-self-excited parallel transductor with alternating-voltage control.

(*a*) Circuit arrangement for parallel-connected control windings.
(*b*) Circuit arrangement for series-connected control windings.
(*c*) Experimental characteristics for a Mumetal-cored transductor. Supply voltage = 20 volts (r.m.s.) 400 c/s, $N_a = 400$, $R = 47$ ohms, $N_c = 200$, $R_c = 1 \cdot 5$ kilohms.

(i) For parallel connection of the control windings.
(ii) For series connection of the control windings.
(iii) For single-element half-wave output.

The change from half-wave to full-wave operation with the parallel-control circuit doubles the right-hand portion of the output curve and leaves the left-hand region almost unchanged. With series-connected control windings the output is about double that for the half-wave circuit for both sides of the characteristic. The voltage induced into the signal circuit with series-connected control windings is double that for the other circuit arrangements, and this explains why the signal input voltage required for minimum output in Fig. 13(*c*) curve (ii) is almost double that for curves (i) and (iii).

Further characteristics for parallel- and series-connected control-winding arrangements to show the effects of phase shifts of the signal voltage, and the changes of shape produced by full-wave rectified bias excitation, are given in Figs. 14(*a*) and 14(*b*). With series-control windings the characteristics can be changed to those for parallel-control if extra windings on the elements are connected in parallel to form a closed-circuit loop. If appreciable resistance is inserted in this loop, or in series with each of the control windings when parallel connected, characteristics intermediate between those of Figs. 14(*a*) and 14(*b*) are obtained.

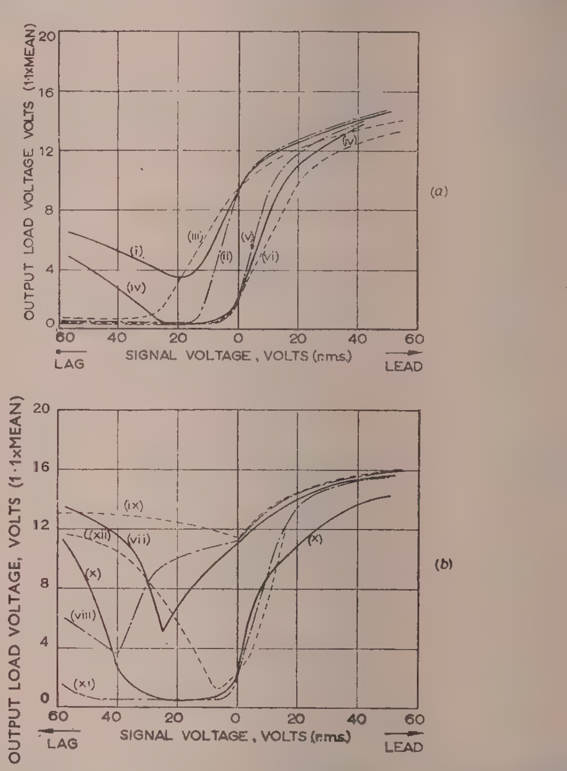


Fig. 14.—Full-wave characteristics for parallel- or series-connected control windings, showing the effects of signal-voltage phase shift (for zero bias and for full-wave rectified-bias condition).

(*a*) Parallel-connected control windings. $N_a = 400$, $N_c = 200$, $N_b = 141$. Supply voltage = 20 volts (r.m.s.) $R_c = 1 \cdot 5$ kilohm, $R = 47$ ohms.

(i) No bias; signal and supply voltages in phase or reverse phase.
(ii) No bias; signal voltage lags supply by $\pi/4$ or leads by $3\pi/4$.
(iii) No bias; signal voltage lags or leads supply by $\pi/2$.
(iv) As (i) with 4·2 mA full-wave rectified bias.
(v) As (ii) with 4·2 mA full-wave rectified bias.
(vi) As (iii) with 4·2 mA full-wave rectified bias.

(*b*) Series-connected control windings.
Test conditions as for (*a*).

(vii) No bias; signal and supply voltages in phase or reverse phase.
(viii) No bias; signal voltage lags supply by $\pi/4$ or leads by $3\pi/4$.
(ix) No bias; signal voltage lags or leads supply by $\pi/2$.
(x) As (vii) with 18·5 mA full-wave rectified bias.
(xi) As (viii) with 18·5 mA full-wave rectified bias.
(xii) As (ix) with 18·5 mA full-wave rectified bias.

Although the curves for parallel-connected control windings are rather better in shape than those for series control, the former arrangement constitutes a damping loop that delays the system response to changes of signal input. The series connection is therefore of greater practical interest.

(5.2) Positive-Feedback Arrangements

In the discussion in Section 4.1 it is shown that, for $R_c = \omega L_c$, the signal ratio V_c/V_p, required for minimum output is $\sqrt{(2)}N_c/N_a$, and from Fig. 7(*a*) curve (ii) the output change for a signal change of this magnitude is about $(1/2\pi)V_p$. The voltage amplification of the curve in terms of the ratio of mean output to r.m.s. signal voltage is therefore $N_a/2\pi N_c$. For 2-element full-wave operation this sensitivity is about doubled, and therefore with a turns ratio of say 20, voltage amplification of 6 or thereabouts (depending on the ratio $R_c/\omega L_c$) may be expected.

It is found that the circuit arrangements used for boost feedback in d.c.-controlled transductors are also applicable to a.c.-controlled units. In Fig. 15(*a*) the circuit diagram shows feedback of the load current through windings arranged to respond to d.c. input. This may be simplified to the boost-feedback form shown in Fig. 15(*b*), as explained elsewhere.[1] Feedback of the load voltage (or the input voltage to the load rectifier), as shown in Fig. 15(*c*), is another method of increasing sensitivity. Alternatively a feedback current proportional to the load voltage may be obtained as shown in Fig. 15(*d*). This maintains isolation between the signal and output circuits, and the shunt feedback effect does not change much with variation of load resistance as occurs with the arrangements shown in Figs. 15(*a*) and 15(*b*).

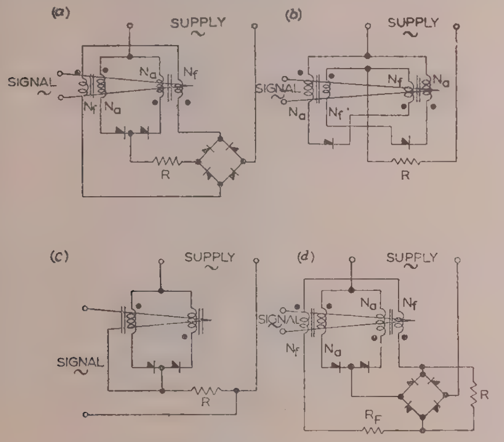


Fig. 15.—Typical feedback arrangements to increase the sensitivity of an a.c. controlled full-wave circuit.

(*a*) Rectified output-current feedback.
(*b*) Boost-excitation arrangement.
(*c*) Positive feedback of the load voltage in series with the signal voltage.
(*d*) Shunt current feedback derived from the load voltage.

With these various circuits it was practical, with the experimental transductor used, to achieve improvements in sensitivity to a.c. control by factors of several times. The positive feedback, of course, has the normal effect of slowing the response to signal-input changes.

(6) PUSH-PULL CIRCUITS WITH A.C. CONTROL

(6.1) Half-Wave Circuit Arrangements

In many transductor applications push-pull forms of characteristic are required having zero output for zero input signal. With two transductor elements such characteristics can be obtained provided that half-wave output action is acceptable.

Typical circuit arrangements are shown in Fig. 16. The elements are connected so that, when no input signal is applied, both cores saturate at the same instant in the appropriate supply voltage half-cycle, but with the signal applied the saturation instants are different and voltage is obtained across the load in the interval between saturation of one core and the other. During this interval, voltage is induced from the main circuit into the control circuit, which therefore has to contain a reasonably large impedance to restrict the resulting current flow.

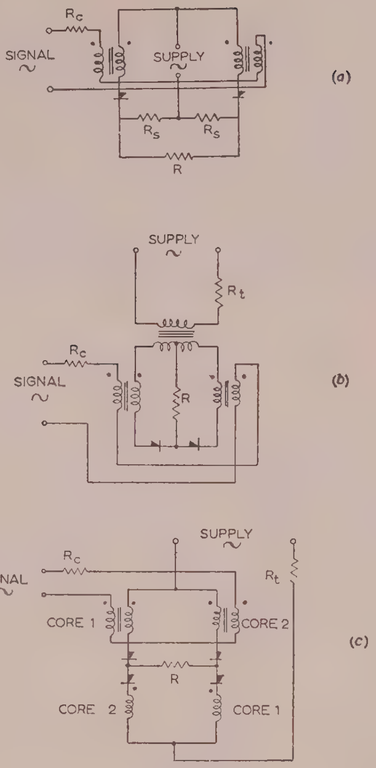


Fig. 16.—Half-wave push-pull transductor arrangements.

(*a*) Circuit described by Whitely and Ludbrook.
(*b*) Arrangement with centre-tap supply transformer.
(*c*) Bridge arrangement used by Lufcy.

In Fig. 16(*a*) the circuit shown is that described by A. L. Whitely and L. C. Ludbrook. Where twin load coils are available, these may be substituted for the resistors R_s, and an overall load is not required.[4]

The arrangement shown in Fig. 16(*b*) is suitable for feeding a single load R which may function on either the d.c. or a.c. components of the output waveform. The resistance R_t shown in the supply-transformer primary circuit serves to limit the current that circulates between the elements when both are saturated.

If each element is provided with two separate main windings the circuit may be rearranged in the bridge form shown in Fig. 16(*c*). This circuit arrangement has been developed by C. W. Lufcy *et al.* for the control of a.c. 2-phase servo motors.[5,6]

Half-wave circuits reduce the number of components to a minimum but have the disadvantage of drawing an unsymmetrical current waveform from the supply source.

(6.2) Performance of a Typical Full-Wave Push-Pull Transductor Amplifier

The main-circuit connections of full-wave push-pull transductors may be in any of the forms normally used for d.c.

control, but for the acceptance of a.c. (supply-frequency) input signals the control windings must be connected in the non-cancellation sense.[7] Of particular interest is the change of sensitivity that occurs when transductors are a.c. instead of d.c. controlled. This effect will be discussed with reference to a particular transductor amplifier operated under both signal conditions so that a direct comparison of the performance difference is possible.

Four identical transductor elements were made on Mumetal clock-spring-type cores $1\frac{1}{4}$ in internal diameter, $1\frac{1}{2}$ in external diameter and $\frac{1}{4}$ in tape width with a main winding of 3 500 turns, a bias winding of 700 turns and a control winding of 175 turns, the wire size for all windings being the same, i.e. 0·0092 in.

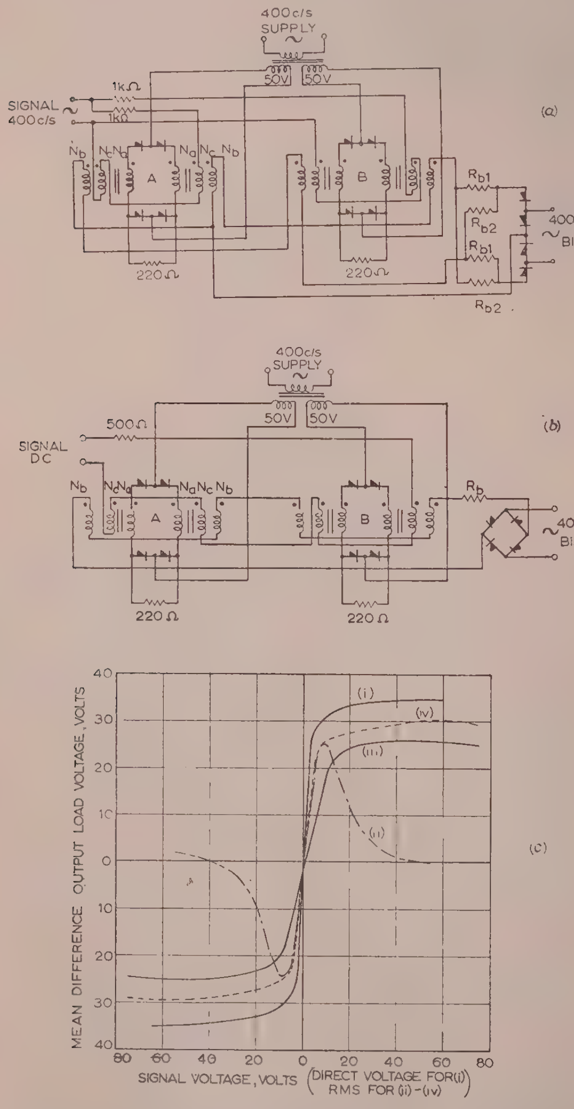


Fig. 17.—Full-wave push-pull transductor amplifiers.

(*a*) A.C. controlled amplifier with divided-bridge bias.
(*b*) Conventional d.c. controlled arrangement.
(*c*) Practical characteristics.

(i) D.C. control (0·4 mA full-wave bias).
(ii) In-phase alternating-voltage control (0·8 mA full-wave bias).
(iii) For $\pi/2$ phase shift of signal.
(iv) In-phase signal with divided-bridge bias (7·5 mA in one half-cycle and 0·4 mA in the other).

These elements were tested as push-pull bridge transductors feeding two output resistors, of 220 ohms each, representing twin load coils for magnetic summing, as shown in Figs. 17(*a*) and 17(*b*). The d.c.-controlled circuit [Fig. 17(*b*)] is quite orthodox and gave the output characteristic of Fig. 17(*c*) curve (i), which has a voltage amplification of 9·6. For a.c. input signals the control windings were rearranged in the non-cancellation sense for each transductor, and the windings for the two transductors were then connected either in series or in parallel. These alternative control arrangements were found to give similar output characteristics provided that the control-circuit resistance in each of the parallel limbs was twice that used for the series-control connection.

In Fig. 17(*a*) the parallel type of control-circuit arrangement (which actually gave slightly better characteristics than the series arrangement) is shown. With normal full-wave bias and an in-phase signal the characteristic obtained was curve (ii) of Fig. 17(*c*), which shows serious swamping effects at quite moderate input-signal amplitudes. With $\pi/2$ phase-shifted signals, curve (iii) was obtained; this does not suffer from swamping effects but is of low sensitivity. Divided-bridge bias, as shown in Fig. 17(*a*), was then substituted for the normal full-wave bridge bias, and with an in-phase signal this gave the characteristic of Fig. 17(*c*) curve (iv). The voltage amplification of this curve is 5 (the turns ratio N_a/N_c being 20), and this is half the sensitivity of the d.c. control condition given. This loss of sensitivity on changing from d.c. to a.c. control is in general agreement with the results obtained theoretically and practically, for single-element and two-element full-wave circuits.

Measurements of the dynamic performance of the push-pull transductor arrangements [Figs. 17(*a*) and 17(*b*)] showed that a step-function change of signal input produced full change of the output-voltage waveforms in from $1\frac{1}{2}$ to 2 cycles of the supply frequency, depending on the instant at which the signal was switched.

(7) CONCLUSIONS

Alternating-voltage signals, of the same frequency as the transductor supply, may be used to control single-element and two-element auto-self-excited transductors provided that the control windings are connected in the non-cancellation sense. The control circuit must contain impedance in series with the windings to restrict the flow of current in this circuit caused by the voltage induced from the main circuit. This induced voltage must be balanced, at least in part, by the signal voltage.

The shape of the output characteristic depends on the phase of the signal voltage in relation to the transductor supply voltage. For in-phase signals, a swamping effect occurs, but a special bias circuit has been devised to overcome this difficulty.

The performance of a typical push-pull 4-element design is examined, and the voltage amplification with alternating-voltage control is 5 (the ratio of the number of main-winding turns to control-winding turns being 20), which is shown to be one-half of that given with direct-voltage control when using the same elements and control-circuit resistance. The response to signal change was complete within $1\frac{1}{2}$–2 cycles of the supply frequency.

The gain of a.c.-controlled transductors may be increased by various boost-feedback techniques similar to those applicable in d.c.-controlled circuits.

(8) ACKNOWLEDGMENTS

The assistance of J. Baranowski with the experimental work and the calculation of theoretical characteristics is gratefully acknowledged.

Acknowledgment is made to the Chief Scientist, Ministry of

Supply, and to the Controller, H.M. Stationery Office, for permission to publish the paper.

(9) REFERENCES

(1) MILNES, A. G., and LAW, T. S.: "Auto-self-excited Transductors and Push-pull Circuit Theory," *Proceedings I.E.E.*, Paper No. 1599 M, April, 1954 (**101**, Part II, p. 643).

(2) SMITH, E. J.: "Determination of Steady State Performance of Self-Saturating Magnetic Amplifiers," *Transactions of the American I.E.E.*, 1950, **69**, p. 1309–17.

(3) MAINE, A. E.: "High Speed Magnetic Amplifiers and Some New Developments," *Electronic Engineering*, 1954, **26**, p. 180.

(4) WHITELY, A. L., and LUDBROOK, L. C.: "Magnetic Amplifier," U.S. Patent 2229952, Jan., 1941.

(5) LUFCY, C. W., SCHMID, A. E., and BARNHART, P. W.: "An Improved Magnetic Servo Amplifier," *Transactions of the American I.E.E.*, 1952, **71**, Part I, p. 28.

(6) LUFCY, C. W., and WOODSON, H. H.: "Design Considerations of the Half-Wave Bridge Magnetic Amplifier," *Communications and Electronics*, July, 1954, **13**, p. 220.

(7) OGLE, H. M.: "The Amplistat and its Application," *General Electric Review*, 1950, **53**, No. 8, p. 41.

(10) APPENDICES

(10.1) Analysis of the Direct-Voltage Control of a Single Element with Positive or Small Negative Signals

The mode of action for this condition, which relates to the high-slope region of the characteristic, is described in Section 2.1 and is represented by the waveforms of Figs. 2(*n*)–2(*r*). The equations for the various circuit conditions are as follows:

Between α and θ the core is saturated, and hence

$$i = \frac{V_p}{R} \sin \omega t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and

$$i_c = \frac{V_c}{R_c} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Between θ and β the core is not saturated and the rectifier conducts. Therefore

$$L\frac{di}{dt} + M\frac{di_c}{dt} + Ri = V_p \sin \omega t \quad . \quad . \quad . \quad . \quad (3)$$

and

$$L_c\frac{di_c}{dt} + M\frac{di}{dt} + R_c i_c = V_c \quad . \quad . \quad . \quad . \quad (4)$$

where $L_c = L(N_c/N_a)^2$ and $M = \sqrt{LL_c}$

Between β and γ the core is still unsaturated, but the rectifier sustains a (reverse) voltage V_r and the main-circuit current is zero. The circuit equations are therefore

$$M\frac{di_c}{dt} + V_r = V_p \sin \omega t \quad . \quad . \quad . \quad . \quad (5)$$

and

$$L_c\frac{di_c}{dt} + R_c i_c = V_c \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Between γ and $2\pi + \alpha$ the circuit conditions are the same as between θ and β, and therefore the basic circuit equations are as eqns. (3) and (4).

Solution of these equations, with suitable boundary conditions, gives for the mean load voltage by integration over a complete cycle

$$V_{av} = \frac{V_p}{2\pi}\left[\frac{\omega L_c}{R_c}(\sin \beta - \sin \gamma) + \cos \gamma - \cos \beta\right] \quad . \quad (7)$$

and gives for the relationships between the angular parameters β, γ and θ

$$\frac{\sin \beta}{\sin \gamma} = \varepsilon^{(\gamma - \beta)/(\omega L_c/R_c)} \quad . \quad . \quad . \quad . \quad . \quad (8)$$

$$\left\{\cos \beta - \left[\frac{RN_c^2}{R_c N_a^2}\left(\frac{\omega L_c}{R_c} + \frac{\omega L}{R}\right) + \frac{R}{\omega L}\right] \sin \beta\right\}\varepsilon^{\beta/[(\omega L_c/R_c)+(\omega L/R)]}$$

$$= \left[\cos \theta + \left(\frac{\omega L_c}{R_c} + \frac{\omega L}{R}\right) \sin \theta\right]\varepsilon^{\theta/[(\omega L_c/R_c)+(\omega L/R)]} \quad . \quad (9)$$

and

$$\sin \theta = \frac{R}{\omega L}\left(1 - \frac{V_c N_a}{V_p N_c}\frac{\omega L_c}{R_c}\right) \quad . \quad . \quad . \quad (10)$$

(10.2) Expressions for Single-Element Behaviour with Large Negative Control Signals (Direct Voltage)

Typical waveforms for the mode of action applicable to the large negative control-signal condition are shown in Figs. 2(*c*) to 2(*g*), and by an analytical treatment similar in character to that indicated in Appendix 10.1 the expressions obtained are

$$V_{av} = \frac{V_p}{2\pi}(1 - \cos \beta) \quad . \quad . \quad . \quad . \quad (11)$$

$$\Bigl\{\left(\frac{\omega L_c}{R_c} + \frac{\omega L}{R}\right) \cos \beta - \sin \beta - \left(\frac{R}{\omega L} + \frac{V_c R N_c}{V_p R_c N_a}\right)$$

$$\left[\left(\frac{\omega L_c}{R_c} + \frac{\omega L}{R}\right)^2 + 1\right]\Bigr\} \times \varepsilon^{\beta/[(\omega L_c/R_c)+(\omega L/R)]}$$

$$= \Bigl[\left(\frac{\omega L_c}{R_c} + \frac{\omega L}{R}\right)^2 \sin \xi$$

$$+ \left(\frac{\omega L_c}{R_c} + \frac{\omega L}{R}\right) \cos \xi\Bigr]\varepsilon^{\xi/[(\omega L_c/R_c)+(\omega L/R)]} \quad . \quad (12)$$

and

$$\sin \xi = -\frac{R}{\omega L}\left[1 + \frac{V_c N_a}{V_p N_c}\left(\frac{\omega L_c}{R_c}\right)\right] \quad . \quad . \quad . \quad (13)$$

(10.3) Alternating-Voltage Control of a Single-Element

In Section 3.2 a description is given of the three modes of action for the single-element circuit behaviour with in-phase and reverse-phase control signals. Only one mode of action is treated here in detail, since the method of analysis is typical of that used in deriving expressions for the other two modes.

Consider, for example, the mode of action represented by the waveforms of Figs. 5(*h*)–5(*m*). Between α and π the core is saturated and the main and control currents are

$$i = (V_p/R) \sin \omega t \quad . \quad . \quad . \quad . \quad . \quad (15)$$

and

$$i_c = (V_c/R_c) \sin \omega t \quad . \quad . \quad . \quad . \quad . \quad (16)$$

During the ensuing period a small magnetizing current flows in the main circuit, but it is convenient to neglect the voltage drop that it produces in the load resistor. This simplifies the analytical treatment and is permissible since the load resistance is low compared with the other circuit impedances. Between π and β the circuit equations are therefore taken as

$$L\frac{di}{dt} + M\frac{di_c}{dt} = V_p \sin \omega t \quad . \quad . \quad . \quad . \quad (17)$$

and

$$L_c\frac{di_c}{dt} + M\frac{di}{dt} + i_c R_c = V_c \sin \omega t \quad . \quad . \quad . \quad (18)$$

During the next period, β to γ, the main-circuit current is zero and the principal-circuit equation is then the same as eqn. (18) with di/dt equal to zero. The remaining period, γ to $2\pi + \alpha$, has the same circuit conditions as between π and β, and eqns. (17) and (18) apply.

The solution of these various circuit equations with the appropriate boundary conditions gives for the analytical expressions

$$V_{av} = \frac{V_p}{\pi}(1 + \cos\alpha) \quad . \quad . \quad . \quad . \quad . \quad (19)$$

$$\frac{V_c}{V_p}\frac{N_a}{N_c} = 1 - \frac{R_c}{\omega L_c}\cot\beta \quad . \quad . \quad . \quad . \quad (20)$$

$$1 + \cos\alpha = \cos\gamma - \cot\beta\sin\gamma \quad . \quad . \quad (21)$$

and $$\frac{V_c}{V_p}\frac{N_a}{N_c}\sin\eta\cos(\gamma - \eta) - \sin\gamma$$
$$= \left[\left(\frac{V_c}{V_p}\frac{N_a}{N_c} - 1\right)\sin\beta - \frac{V_c}{V_p}\frac{N_a}{N_c}\cos\eta\sin(\beta - \eta)\right]\varepsilon^{(\beta-\gamma)/\left(\frac{\omega L_c}{R_c}\right)}$$
$$. \quad . \quad . \quad . \quad (22)$$

where $$\eta = \arctan(\omega L_c/R_c) \quad . \quad . \quad . \quad . \quad . \quad (23)$$

621.385.16.029.65

Monograph No. 142 R
Aug. 1955

A MILLIMETRE-WAVE MAGNETRON

By J. R. M. VAUGHAN, B.A.

(*The paper was first received* 14*th January, and in revised form* 7*th April*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in August*, 1955.)

SUMMARY

The construction and performance of a pulsed-millimetre-wave magnetron are described. The performance is compared with that of a "parent" valve, using Slater's scaling laws, and with theoretical standards. Some problems encountered in designing the valve and in holding close tolerances on the performance are discussed. The output coupling is analysed in some detail, including the effects of possible constructional errors. Test procedure is briefly described.

LIST OF PRINCIPAL SYMBOLS

V, I, B = Operating voltage, kV, current, amp, magnetic flux density, kG, respectively.

V_0, I_0, B_0 = Hartree's and Slater's "characteristic" voltage, current, magnetic flux density, respectively.

V_R, I_R, B_R = "Reduced" voltage, current, magnetic flux density (expressed as ratios, e.g. $V_R = V/V_0$).

V_1 = Voltage at standard operating point (10 amp, 10 kG).

λ_π, f_0 = π-mode wavelength and frequency, respectively.

f_p = Pulling figure.

N, n, γ = Number of resonators, mode number, Hartree harmonic mode number ($\gamma = n \pm mN/2$, m integral), respectively.

Q_E = External or coupled Q-factor.

P = Perimeter of cross-section of two consecutive resonators.

r_1 = Circuit ratio (ratio of radial depths of resonators).

r_2 = Ratio of vane thickness to resonator gap at mouth.

r_a = Radius of anode.

r_c = Radius of cathode.

$Z_1, Z_2, Z_3, l_1, l_2, l_3$ = Characteristic impedances and lengths of resonator (when considered as a plain quarter-wave line), transformer, and output waveguide up to the "puller," respectively.

Z_p = Equivalent series impedance of puller.

$Z_c = R_c + jX_c$ = Impedance at junction of transformer with output guide, outwards.

Y_r = Admittance of a resonator (Y_{rl}, for long resonator; Y_{rs} for short resonator).

$Y_A = G_A + jB_A$ = Admittance at mouth of resonator to which output is connected.

Y_B = Admittance at back of resonator to which output is connected, outwards.

$a, b, b_l, b_s, b_0, t, \psi, l_a$ = Resonator dimensions as shown in Fig. 7.

κ, κ' = Small errors.

p, q, r, s = Constants determined from resonator dimensions.

σ = Detuning fraction.

β = Phase-change coefficient.

(1) INTRODUCTION

The paper describes the application of the known principles of magnetron design[1, 2] to the production of a millimetre-wave valve, under a rigid specification. Research was undertaken to a limited extent only and the paper is therefore descriptive rather than analytical. A number of Collins's formulae are quoted[2] for comparison with the experimental results, and an extended analysis of the pulling figure is given, since this parameter proved difficult to measure and specify.

The valve is a fixed tuned magnetron for high-power pulse operation at a wavelength of 8·6 mm. It was first made, as the VX9005, by the Services Electronics Research Laboratory, in 1947. Basically, it was a scaled version of the 3J31 magnetron (see p. 786 of Reference 2) although very different externally. The general appearance with cover removed to show the integral magnets is shown in Fig. 1. The designs developed by S.E.R.L. were made available to the author in 1948, and work continued in both laboratories for several years, with frequent interchanges of information. A major part of the results given in the paper are thus due to the work of S.E.R.L. The valve to be described is known as VX5027 and is interchangeable with the final design of type VX9005, although there are some internal differences.

(2) SPECIFICATION AND PERFORMANCE

The basic power and wavelength specifications were decided at the outset, and tolerances and additional restrictions were added as the experience of maker and user increased; the pulling figure, for example, was largely governed by the amount of electronic tuning obtainable in the associated klystron, described in a companion paper.[3] While all valves were tested, as far as possible, at the specification figures, some were run under a much wider range of conditions. Peak powers up to 50 kW and mean powers up to 40 watts have been obtained, but at the expense of cathode life and only by virtue of special cooling for the output window. A general picture of the performance required and achieved is given by the following figures:

Input

Cathode voltage (negative, pulsed)	specified max.	: 16 kV
	actual average	: 13 kV
Cathode current (peak)	nominal	: 10 amp
operation normally satisfactory	from min.	: 6 amp
	to max.	: 14 amp
Cathode current (mean)	specified	: 4 mA
Pulse duration	specified	: 0·2 μsec
Pulse repetition frequency	specified	: 2 000 pulses/sec
Duty cycle		: 0·0004

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Vaughan is with E.M.I. Research Laboratories, Ltd.

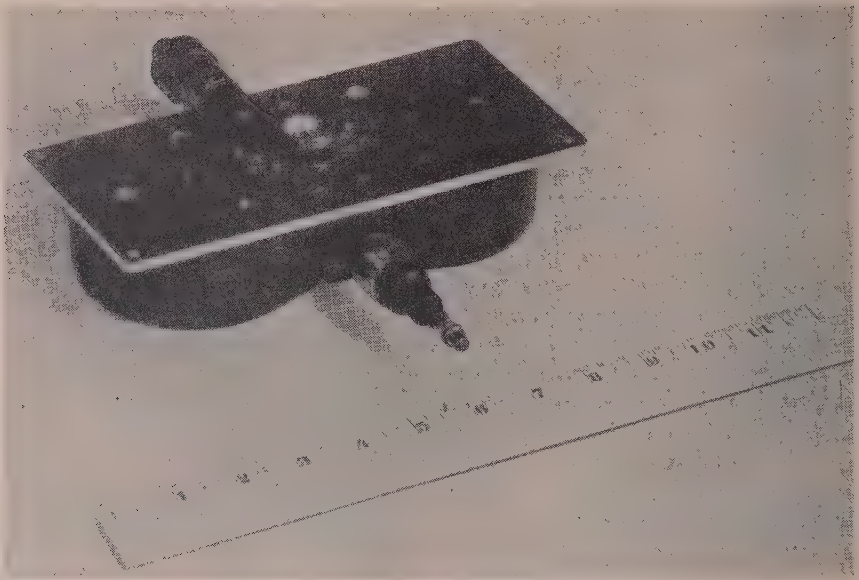


Fig. 1.—The VX5027 magnetron.

The heater terminal is nearest the camera, and beyond it are the cathode terminal, cathode insulator glass and Kovar tube, leading into the pole piece. The output coupler projects through the centre of the clamping plate. The ten small holes close to it are air cooling channels; the two larger holes are accurately spaced with respect to the coupler, and mate with spigots on the equipment. The magnets are protected by a Bakelite cover, not shown. The scale is in inches.

Output

Power output (pulse) (at specified input)	specified min.	15 kW
	actual average	20 kW
(mean) (at specified input)	specified min.	6 watts
	actual average	8 watts
Wavelength	specified max.	8·69 mm
	specified min.	8·51 mm
Pulling figure	specified max.	45 Mc/s
	actual average	32 Mc/s
Bandwidth (¼ power points)	specified max.	20 Mc/s
	actual average	10 Mc/s

(3) ELECTRONIC DESIGN

Initially the design of the resonator and interaction-space shapes was a direct scaling down of the 3J31. Reference to Slater's scaling equations (Reference 2, p. 414 et seq.) shows that the characteristic voltage V_0 and current I_0 are unchanged, and that the characteristic magnetic flux density B_0 is increased in inverse proportion to the wavelength. The characteristic power P_0 $(= I_0V_0)$ and conductance G_0 $(= I_0/V_0)$ are also unchanged. After some minor adjustments and correction of a misprint in Collins's value for P_0, the values for the VX5027 were:

$$V_0 = 3{\cdot}29\,\text{kV}$$
$$I_0 = 20\,\text{amp}$$
$$B_0 = 4{\cdot}27\,\text{kG}$$
$$P_0 = 66\,\text{kW}$$
$$G_0 = 0{\cdot}0061\,\text{mho}$$

If the actual operating voltage V, current I and flux density B are expressed as "reduced variables" V_R, I_R, B_R in terms of V_0, I_0, B_0 (i.e. $V_R = V/V_0$, etc.), the reduced variables may be plotted as a reduced performance chart, which should be closely similar to that of the parent valve.

A stylized average reduced performance chart for the VX5027 is shown in Fig. 3. It is based on the performances of 30 valves, and is stylized in the sense that the best values of the constants were determined for each valve, the results averaged, and the chart plotted from these figures; curvature of the constant-fiel lines is thus eliminated, which is unrealistic but assists com parison with the 3J31 reduced performance chart (fro Reference 2, p. 431) shown superimposed. Both of these ma be compared with a partly theoretical, partly empirical, reduce performance chart based on Hartree's equations (Reference p. 340). These equations, rearranged and corrected for misprint, are as follows:

$$V - 2(V_0/B_0)B + V_0 = 0 \quad . \quad . \quad . \quad . \quad ($$

$$V_0 = 2(m/e)(\pi c r_a^2)(\gamma\lambda)^{-2} \quad . \quad . \quad . \quad . \quad ($$

and

$$B_0 = \frac{4\pi c r_a^2}{r_a^2 - r_c^2}\frac{m}{e}\frac{1}{\gamma\lambda} \quad . \quad . \quad . \quad . \quad ($$

in which the symbols have the usual meaning.

Collins (Reference 2, p. 340) uses V_γ and B_γ to make clear t dependence on γ for various modes. Here we are only concern with the single value $\gamma = N/2 = 9$. Substituting the VX50 dimensions, we have

$$V - 1{\cdot}52B + 3{\cdot}29 = 0 \quad . \quad . \quad .$$

expressed in kilovolts and kilogauss, or

$$V_R - 2B_R + 1 = 0 \quad . \quad . \quad . \quad . \quad (4$$

expressed in reduced units.

Eqn. (4) relates V to B at the onset of oscillations; the furt variation of V as increasing current flows is not predicted theory, but examination of some representative performa charts shows that dV/dI is nearly constant over a large part the charts; the regions excluded are those near zero curre and those at high magnetic flux densities. We may theref add to eqn. (4) an empirical term $-kI$ in which $k = dV/dI$. study of Clogston's reduced performance charts (Reference

p. 419–434) leads to an estimate of 0·11 for k in the Q-band egion. We thus obtain the theoretical-empirical relations

$$V - 1{\cdot}52B - 0{\cdot}11I + 3{\cdot}29 = 0 \quad . \quad . \quad . \quad (5)$$

$$V_R - 2{\cdot}00B_R - 0{\cdot}68I_R + 1{\cdot}00 = 0 \quad . \quad . \quad (5a)$$

vhich are plotted in Figs. 2 and 3 respectively.

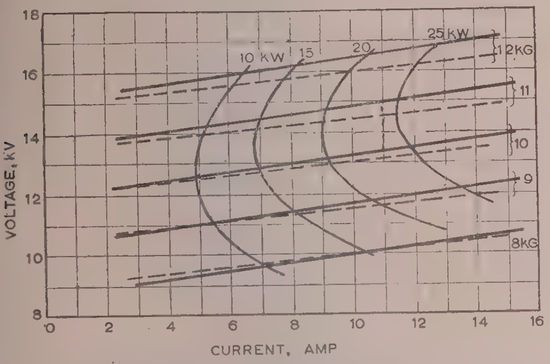


Fig. 2.—Performance chart.

———— Stylized average actual performance.
— — — — Theoretical-empirical performance.

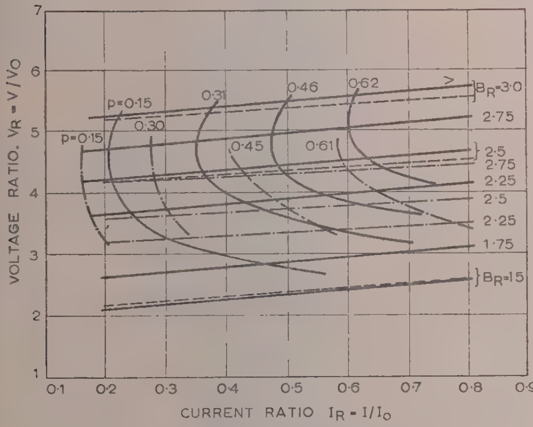


Fig. 3.—Reduced performance chart.

———— Stylized average reduced performance.
— — — — Theoretical-empirical reduced performance.
— - — - — Reduced performance of 3J31 magnetron for comparison.

The equations describing the stylized average actual and educed performance charts are as follows:

$$V - 1{\cdot}59B - 0{\cdot}127I + 4{\cdot}0 \ = 0 \text{ (Fig. 2)} \quad . \quad . \quad (6)$$

$$V_R - 2{\cdot}09B_R - 0{\cdot}78I_R + 1{\cdot}23 = 0 \text{ (Fig. 3)} \quad . \quad (6a)$$

The scatter of values of the coefficients is shown in Table 1, vhich also includes the voltage V_1 at the standard operating oint (10 amp, 10 kG), and the pulling figure f_p, for which a heoretical value is derived later. To those accustomed to naking measurements at Q-band frequencies, perhaps only the ulling-figure variations will seem excessive (see Section 8); the thers reveal no great inconsistencies. The high apparent value or V_0 is due partly to the high value of dV/dI.

The reduced conductance g is 0·145.

Table 1

Comparative Performance of 30 Valves

Quantity	Mean	Octiles*	Extremes	Theoretical
dV/dB	1·59	1·4 1·75	1·35 1·85	1·52
dV/dI	0·127	0·11 0·14	0·08 0·26	
V_0	4·0	3·0 5·0	2·1 6·0	3·29
V_1	13·0	12·4 13·6	11·4 14·7	12·4
f_p	32	20 40	9 61	42†

* 75% of valves fall between the octiles.
† For standard 0·012 in coupling transformer.

(4) CONSTRUCTION

The following considerations determine the general layout:

(*a*) Owing to the high magnetic flux density required, the pole-pieces must be built into the valve, close to the interaction space (between cathode and anode) at each end.

(*b*) The centring of the cathode with respect to the anode is critical, while the axial position is less important; the thermal expansion of the cathode support members during operation precludes an unsymmetrical radial mounting, while the space between anode and either pole-piece is not enough for a symmetrical radial mounting. The cathode must therefore be supported axially through a hole in a pole-piece.

(*c*) The length of insulator required to withstand the pulsed cathode voltage is much larger than the radius of the permissible hole in the pole-piece. The cathode must therefore be supported on a rather long stem passing right through the pole-piece to a large insulator beyond. This system has the obvious disadvantage that, under conditions of vibration, the support may be inadequate, and this has not yet been eliminated.

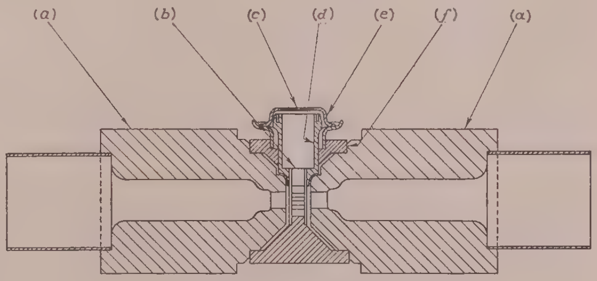


Fig. 4.—Body assembly.

(*a*)—Pole-piece.
(*b*)—Transformer.
(*c*)—Window.
(*d*)—Waveguide output.
(*e*)—Window cup.
(*f*)—Anode block.

The body assembly (Fig. 4) comprises the copper anode containing the 18 resonant cavities, pole-pieces and short lengths of Kovar tube arranged along the main axis, and the output coupling transformer, waveguide and window, lying on a transverse axis. The parts are assembled in a jig with the necessary rings of copper–silver eutectic solder, and covered by a glass vessel filled with hydrogen. The pole-pieces are heated by radio-frequency induction, and all the brazes are completed in one operation, no flux being required. This method is not only

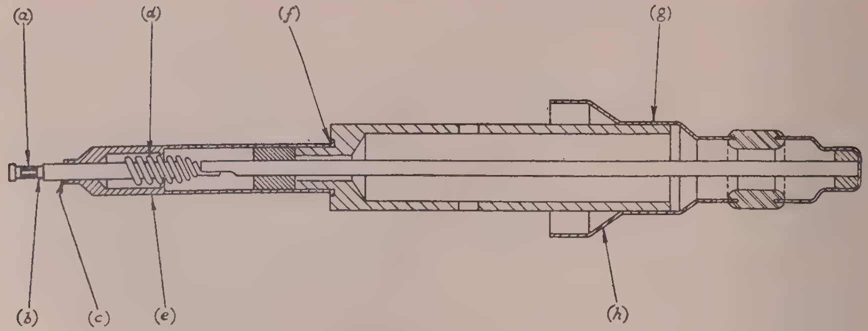


Fig. 5.—Cathode assembly.

(*a*)—Nickel mush.
(*b*)—Tungsten cathode spindle.
(*c*)—Nickel-palladium solder braze.
(*d*)—Heater.
(*e*)—Nimonic-alloy tube.
(*f*)—Copper-gold solder braze.
(*g*)—Copper-silver solder braze.
(*h*)—Kovar hat.

much quicker than furnace brazing but is less liable to cause distortion of the output transformer.

The cathode assembly (Fig. 5) is entirely symmetrical axially, in order to minimize radial movements during heating up. The active surface, which is considered in detail in Section 6, is carried on a tungsten core, brazed with nickel–palladium solder to a Nimonic-alloy support; Nimonic alloy, developed for gas-turbine blades, is chosen for its low creep at high temperatures. This in turn is brazed to a steel tube, and that to a Kovar hat adapted to seal both to the heater insulator and to the main cathode insulator. This assembly is also brazed by radio-frequency induction under hydrogen, but the brazes are done sequentially.

The insertion, axial positioning and centring of the cathode assembly, without damage to the delicate vanes of the anode, is a major problem. The jig used (Fig. 6) provides micrometer movements in three perpendicular directions, while the relative positions of the parts are being observed through two 25 × tool-room microscopes. Errors of parallelism are corrected by a ball-and-socket clamp, and independent centring adjustments are provided for the axial microscope and for the coil of the induction heater, by which the Kovar parts are sealed to the ends of the glass cathode insulator, the axial and radial adjustments being completed while the glass is soft.

Pumping is controlled by reference to the pressure gauge rather than the clock; baking at 450° C and cathode activation are each continued until the pressure falls to 10^{-6} mm Hg. The pressure is measured with an ionization gauge placed in the pumping tube so that gas travelling from the valve to the pump must pass through it. Errors due to the pumping action of the gauge itself are thus minimized.

The permanent magnets and some other external components are not added until after the valve has been tested in an electromagnet.

Some tests with gear capable of flexing the Kovar cathode-hat while the valve is running have indicated an accuracy of centring that may be expressed as over 90%, in the sense that not more than 10% increase of output power can be obtained by any further adjustment.

The central hole and 18 surrounding resonator cavities of the anode are formed by hot hobbing. A hardened steel tool is ground to the required shape and forced into an oxygen-free high-conductivity copper blank at 600° C. The process is one of plastic flow, not of cutting, and wear on the hobs is negligible; failure, when it occurs, is always irreparable. To enable th block to be faced off (after extraction of the hob) withou damage to the vanes, the hole is filled with Diakon, under moulding pressure of 4000 lb/in^2. Moulded thus, it is sul

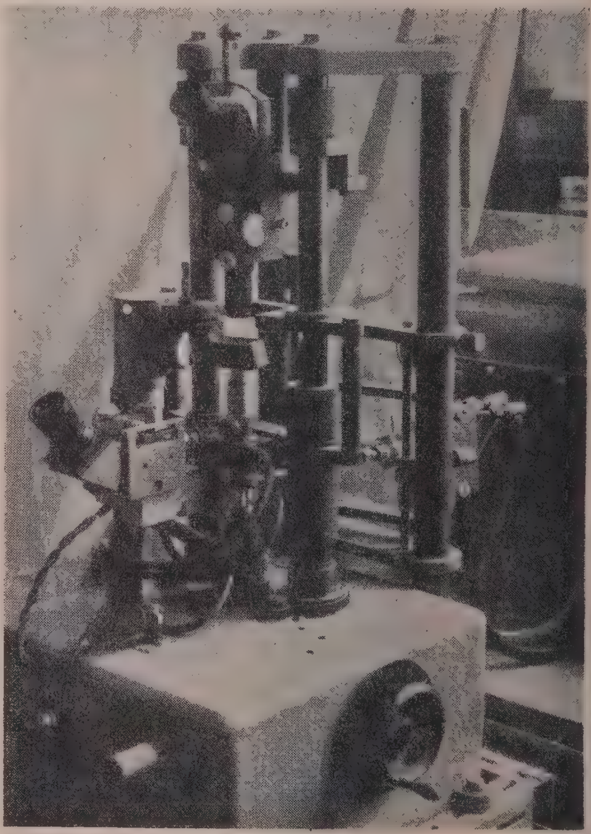


Fig. 6.—Cathode centring jig.

A water-cooled clamp, holding the body, is moved vertically by the handwheel c the right. Below it the cathode is held by a water-cooled pneumatic collet adjustab horizontally and for tilt. The eddy-current heater coil is seen in position for sealir the cathode insulator glass to the body. The centring and axial adjustments a observed through the two microscopes.

tantially as hard as the copper, and the machining can be arried out with a minimum of burrs.

The H-section hole in the output transformer and the rectngular hole in the waveguide are also hobbed.

The hobs are ground automatically on machines designed y H. E. Holman for the purpose.[4] A light cut (usually ·0002 in) is taken on each slot or face in turn, the work being ndexed after each cut, and the wheel fed after each revolution. n this way, uneven hardness of the material does not cause nsymmetrical wear of the wheel, and deep, narrow slots can e ground with uniform accuracy. Hobs are normally made rom HPSH (an 18% tungsten tool-steel), hardened to 700–50 Vickers Diamond Point, or from JC20 steel, hardened to 00–550 V.D.P. A small chamfer (about 0·001 in) is necessary n all sharp exterior angles, to prevent stress concentrations from ausing cracks and splintering.

(5) RESONATOR SYSTEM

The resonator system (Fig. 7) is an 18-vane, sectoral-cavity, ising-sun, open-ended block, with circuit ratio $r_1 = 1 \cdot 8$. The imensions shown are those finally evolved for $\lambda_\pi = 8 \cdot 6$ mm.

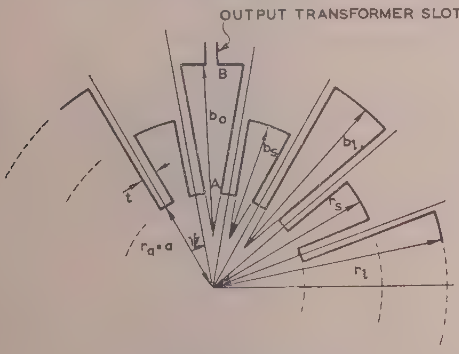


Fig. 7.—Resonator system.

$r_l = 0 \cdot 140$ in.
$r_s = 0 \cdot 1017$ in.
$r_a = a = 0 \cdot 055$ in.
$l_a = 0 \cdot 104$ in.
$\psi = 20°$.
$t = 0 \cdot 0095$ in.

he dimensions found by S.E.R.L. for the same wavelength are reater by about 0·0025 in for r_l and 0·0015 in for r_s. This pparent discrepancy is entirely accounted for by the fact that he vane tips of the S.E.R.L. design were made almost semiircular in cross-section, while those for the VX5027 were made s nearly square as possible, the corner radius being about ·0015 in. The difference represents more than twice the permissible tolerance on wavelength, and illustrates the great care eeded in grinding the hobs. The control of shape at the bottom f a slot 0·085 in deep ×0·010 in wide is by no means easy. It s evidently necessary that whatever profile is used it should be losely uniform round the 18 vanes. Apart from the question of avelength, no difference of performance has been noticed that ight be attributed to the choice of round or square tips.

(5.1) Mode Separation

With $r_1 = 1 \cdot 8$ and $D_a = 0 \cdot 325\lambda_\pi$, the separation of the mode from the π mode is 5% downwards, and that of the mode is over 25% upwards. Direct interference with the mode is thus negligible. Some might be expected from reverse omponents ($m = -1$) of the long-wavelength modes 1 to 4, ut none has been observed. The short anode length $l_a = 0 \cdot 31\lambda_\pi = 0 \cdot 104$ in) precludes interference from the axial armonics of these modes.

(5.2) Determination of Wavelength

The wavelength can be determined in theory from a solution of Maxwell's equations, which are separable for the sectoral cavity. In practice the unknown admittance of the open-ended central cavity is the major factor in causing the result to be about 10% low. The solution is quoted in Section 8 [eqn. (20)] in connection with the calculation of the pulling figure.

A more useful equation is the empirical perimeter formula given on p. 479 of Reference 2:

$$\lambda_\pi = P[1 \cdot 03 - 0 \cdot 06(r_1 - 1 \cdot 8) + 0 \cdot 05(r_2 - 1 \cdot 5)] \quad . \quad (7)$$

where P is the perimeter of cross-section of two consecutive cavities, and r_2 is the copper/space ratio at the resonator mouths. Eight sets of dimensions, involving 31 valves, were checked against this formula. After allowing 0·5% for the residual vane-tip curvature, the formula was found to be within 0·5% of the mean observed wavelength. The scatter of wavelengths of valves made from any one hob has a standard deviation of 0·035 mm (0·4%), most of which is due to small variations in the spacing between anode and pole-pieces.

(6) CATHODES

The working area of the cathode, effectively determined by the anode dimensions, is 0·13 cm^2. Hence the peak-current rating of 10 amp represents 75 amp/cm^2, and at least 100 amp/cm^2 is required on test. This is not obtainable as primary emission from any cathode with useful life, but the back bombardment of the cathode by electrons accelerated inwards by oscillations in the space charge, followed by the emission of secondary electrons, provides a means of building up large currents from a small primary emission. The same process accelerates some other electrons outwards to strike the anode, although the valve is nominally cut off by the magnetic field, so that the h.t. supply provides the energy required. The electrons reaching the anode have velocities less than that corresponding to the applied voltage, and the balance of the energy supplied appears in part as excess energy of the electrons returned to the cathode, which thus becomes heated. The build-up of current occurs in less than 10^{-8} sec; the period is difficult to observe accurately, owing to the large capacitive currents generated by the high rate of rise of voltage.

The back bombardment is therefore essential in the operation of the valve, but it has additional effects which are harmful: it tends to remove the coating bodily from the cathode, and to raise the temperature, thus increasing the loss of coating by evaporation. The evaporation can be offset, within limits, by reducing the heater input while the valve is running. The usual distinction between space-charge-limited and temperature-limited emission is obscured in this case by the secondary emission build-up; the effect of reducing the heater input is not to reduce the anode current at all until a point is reached at which the primary emission is not sufficient to start a pulse. The valve then ceases to oscillate and the anode current falls abruptly to zero. In some cases the valve may continue to run with zero heater input; although at first sight convenient, this is unsatisfactory since it implies that the back bombardment is abnormally high, and the life will be correspondingly short.

The first cathodes used in the VX5027 consisted of a nickel rod coated with barium and strontium oxides in the usual way. The coating was lost in a few hours. Grooves were then turned in the rod and filled with coating, which increased the life to a few tens of hours, and the nickel rod was changed to Nimonic alloy to obtain better hot strength. The next step was to sinter nickel powder into a wide recess, 0·010 in deep, to form a porous matrix, which was impregnated with the oxide coating by soaking. The life was increased by this means to 250–500 h, although it is probable that some of the increase should be attributed to reduction of back bombardment by the better methods of cathode centring that had come into use. The nickel particle size does not appear to be critical. S.E.R.L. used 250–300-mesh powder, sintered in cracker gas. This coarse powder would not sinter to Nimonic alloy in cylinder hydrogen, owing to the rapid oxidation of the chromium content, and cracker gas was not conveniently available, so much finer (5 μ) powder was used, without apparent effect on the cathode life. Finally the Nimonic-alloy core was changed to tungsten, and lives of 500–2 000 h were obtained.

Some other cathodes, such as platinum-plated types with grooves beyond the ends of the interaction space, and nickel–tungsten alloy loaded with barium beryllonate, were also tried, but were less successful. The successful cathodes were developed at S.E.R.L.

The heater is of the "soldering-iron" type, set back inside the pole-piece to protect it from the magnetic field. Apart from the smallness of the space that would be available if the active part of the cathode were made hollow, a helical heater in the centre of the valve would (at normal warming-up current) suffer radial stresses of the order of 50 g, alternating at the frequency of supply.

(7) MAGNETIC FIELD

The valve operates satisfactorily at flux densities in the range 8–11·5 kG, the best efficiency being at about 10·5 kG; above this the efficiency falls as the cyclotron resonance point (14·3 kG) is approached (see Reference 2, p. 437). Valves were occasionally operated above this point, at about 16 kG, but not without great detriment to the cathode.

The original pole-pieces, cut away to accommodate the slot transformer, were found to give an asymmetric field. Curve A in Fig. 8 shows the variations along the transverse axis through the centre of the output waveguide. After the change to an H transformer, almost all the cut-away was eliminated, and curve B was obtained. These curves were plotted by measuring the Hall effect in a piece of germanium, 0·030 × 0·030 × 0·015 in, moved on a probe along the transverse axis of a hollow dummy. They were reproducible within 1% as regards shape, but the absolute values were doubtful to about 5% owing to uncertainty about the Hall coefficient.

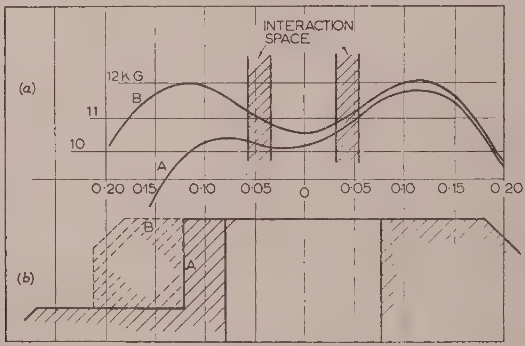


Fig. 8.—Variation of axial magnetic flux density along transverse axis.

(a) Measured flux, showing effect of cut-away pole-piece. Horizontal scale is in inches.
(b) Transverse cross-sections of noses of corresponding pole-pieces, to same horizontal scale.

The presence of a minimum at the centre of the valve implies that the magnetic equipotentials are concave towards this point. Hence the orthogonal lines of force are convex towards t centre; i.e. the field in an axial plane is of the pin-cushi type in the region of interest. This tends to push electrons c of the interaction space and so reduces efficiency, but it almost unavoidable with an axially mounted cathode. T magnetic circuit is thus very inefficient, the peak field lyi entirely outside the interaction space, and only about 5% the total flux being used. The gap is 0·174 in.

The flux density of 10·5 kG was provided by 5 500 ampere-tur in an electromagnet for testing, or by a pair of horseshoe magn designed by A. J. Tyrrell, and weighing 3¼ lb each.

(8) OUTPUT COUPLING

Various tests have shown that loading applied to a sin cavity is distributed uniformly by the mutual coupling of t cavities. The output is therefore taken from the back of a lo cavity, through a quarter-wave transformer to a standa Q-band waveguide, in which is placed a vacuum windo (Section 9). As the coupling was often found to differ grea from the first-order theory (see Reference 2, Section 5.3), closer analysis was made of the effects of the long cavity, a of some possible geometrical errors. The details are tediou but an outline of the method and results is given below and Sections 13.1 and 13.2.

Physically, the transformer is a quarter wavelength of gui of low characteristic impedance; as first designed, it was plain rectangular guide, 0·244 × 0·015 in in cross-secti ($Z_0 \simeq$ 30 ohms), but this required cut-aways in the pole-piec (Section 7). An H-section guide was later adopted, the wid being thus reduced to 0·140 in, and the girder shape maki hot hobbing possible without buckling the hob.

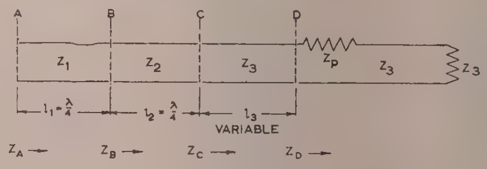


Fig. 9.—Output system.

Z_1 = Characteristic impedance of resonator.
Z_2 = Characteristic impedance of transformer.
Z_3 = Characteristic impedance of output waveguide and matched load
Z_p = Impedance of puller.
Z_A = Impedance at A towards the right.
Z_B = Impedance at B towards the right.
Z_C = Impedance at C towards the right.
Z_D = Impedance at D towards the right.

The degree of coupling between the resonant system and output waveguide is normally expressed in terms of the pull figure, f_p. This is the total frequency shift of the magnet when operating into a load whose v.s.w.r. is 1·5 : 1, which is varied through all phases. It can thus be measu directly on an oscillating magnetron, with the aid of a calibra variable mismatch unit (puller) in a matched guide. In prac the measurement is seldom very accurate; the puller m generally be calibrated with powers of the order of milliwa using a local oscillator, and the calibration is then assume hold with powers of the order of kilowatts, and it is difficul remove all other reflections from the waveguide while absorbing the magnetron power satisfactorily.

Values for f_p may also be found by calculation from the dim sions of the magnetron, and by low-level Q-factor measureme These are known as "cold" pulling figures, since no accou taken of the electron stream present in the oscillating valve.

The Q-factor measurement method, devised by Lawso described in Reference 2, pp. 714–717, but the graph given t

Fig. 18.15) is not applicable when the minimum shift is retro-rade near resonance. The modification required is clear on nspection of a Smith chart. If f_0 is the unperturbed frequency nd Q_E the external Q-factor, obtained from observations of oaded Q-factor and resonance v.s.w.r., the cold pulling figure is asily shown to be

$$f_p = \frac{5}{12}\frac{f_0}{Q_E} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

In practice, there is usually reasonable agreement (generally vithin 25%) between cold pulling figures so measured and the lirectly observed hot pulling figures. But in a substantial iumber of cases with the VX5027, both were reduced by a actor of up to 5 from the value expected from the magnetron limensions.

The theoretical value of the cold pulling figure is obtained in tages. First, the cavities and transformer are all assumed to be imple quarter-wave lines, and the load positions for maximum requency-shift are found. Next, this is corrected for the sectoral avity shape and the alternation of sizes, assuming the same oad positions to apply. Finally, a correction is made for the I transformer. The analysis is outlined in Section 13.1. For a ransformer gap of 0·015 in as originally used, the result is

$$f_p = 66\,\mathrm{Mc/s}$$

This is greater than the specification limit, and as some alves were found to approach it (it was seldom exceeded), the ransformer slot size was reduced to 0·012 in; the square-law ariation of f_p then gave 42 Mc/s.

It was found that most good valves had hot pulling figures of bout 75% of this theoretical value, with some as low as 20% nd others up to 140%. The high values had no serious conse-uence, since such valves generally had high power output, and here were uses for them in which the pulling figure was unim-ortant. But the low values meant a serious loss of valves, and equired close investigation. Careful examination revealed no eometrical errors or physical flaws, and tests showed that ieither the window nor the cathode was responsible. A definite orrelation was found between reduction of pulling figure and small increase of wavelength, but this led to no further lucidation of the problem.

The effect of errors in the transformer length was analysed Section 13.2); it was shown that no serious change of pulling igure would result from any reasonable error, but there emerged n explanation of the S-shaped distortion of the Rieke diagram ften observed in practice [Figs. 10(*a*) and 10(*b*)].

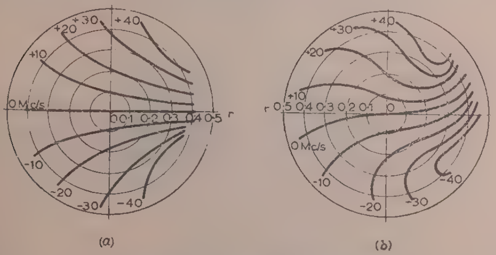


Fig. 10.—Rieke diagrams.

(*a*) Undistorted.
(*b*) Distorted by incorrect transformer length.

It was also thought possible that the transformer might be ressed too hard against the back of the output cavity during ssembly, so that the vanes would be pressed inwards and apart, ecreasing the cavity angle and increasing the gap at the mouth. nalysis similar to that in Section 13.1 (ii) showed that a small percentage decrease κ' in ψ (cavity angle) would produce an increase in f_p of only $2\kappa'$ per cent.

The problem thus remains unsolved. Fortunately valves with abnormally low pulling figure have not been encountered recently.

(9) WINDOW

The output window is a circular metal iris closed with low-loss glass, between, and spaced from, sections of standard rectangular guide; the central parts of the broad faces are flared at the window to reduce the voltage gradients at the discontinuity, and propagation radially outwards along the gap (which is necessary for thermal-expansion matching) is interrupted by quarter-wave chokes. The complete assembly of flares, chokes, iris and glass must be designed as one unit, to obtain a low v.s.w.r. It is found that there is not, as in simple windows, a set of combinations of iris diameter and glass thickness giving a v.s.w.r. of unity, but one pair only, for fixed values of the other parameters. It is essential that the glass disc should be sunk into the plane of the iris, otherwise the phase difference of the glass and iris reflections makes a match impossible without the use of further elements. Ceramic windows, sitting on the iris, gave standing-wave ratios of not better than 1·3, although this could have been improved by a symmetrical double-iris construction. With glass, a value of 1·03 was generally obtained. The glass first used was Kodial, which "sucked in" at a mean power output (matched load) of 16–20 watts. Corning 7070 glass was used subsequently and withstood 30–35 watts. However, Kodial windows fitted in the waveguide at a distance from the magnetron, with vacuum on one side and the same ambient temperature as on the magnetron, have passed over 40 watts safely. Less than half the power absorbed is therefore due to dielectric loss; the question whether the balance is due to heat radiation from the cathode, to X-rays, or to electron or ion bombardment is still being studied.

When suck-in occurs, the puncture is a neat ellipse, about 2 × 1 mm, with the major axis parallel to the electric field, in the centre of the glass.

(10) TESTING

Each valve is aged and tested individually to a level about 50% above the rated input. The high voltage gradient between cathode and anode (up to 30 kV/mm) causes some initial sparking, and the valve is run for several hours, to ensure that this has ceased, before the performance is charted. The output power is absorbed in a continuous-flow water calorimeter consisting of a silica tube passing obliquely through the waveguide. The temperature rise is measured by four thermistors arranged in an unbalanced bridge. The rate of flow having been correctly set, the mean power is read on a panel meter calibrated directly in watts.

About 1% of the power passes the calorimeter and is fed through a low-level variable attenuator to the spectrum analyser. This consists of a cylindrical cavity wavemeter, operated in the H_{01} mode, with a small oscillating plunger at the centre of one end face. The wavemeter is manually tuned by a micrometer plunger at the opposite end, while the small plunger sweeps the tuning rapidly over ±50 Mc/s in synchronism with the X-sweep of an oscillograph. The wavemeter response is fed to the Y-plates, and spectrum is thus displayed. The wavemeter may also resonate in the E_{01}, E_{11}, H_{11}, H_{21} and H_{31} modes, but all except the last are largely suppressed by the placing of the coupling holes and the use of non-contact plungers. The H_{31} response is strong, but since the energy is concentrated round the periphery of the cylinder the response is not affected by the small central oscillating plunger; the oscillograph thus shows a uniform response which cannot be confused with the normal peaked spectrum.

The pulling figure is read directly from the spectrometer by setting the puller to a v.s.w.r. of 1·5 and moving it along the guide, the limits of shift of the spectrum peak being noted.

Three oscillographs are used to observe the voltage/time, current/time and voltage/current curves. Owing to the high rate of rise of voltage (500–600 kV/microsec) and high input capacitance (about 25μμF including leads), the charging current is 12–15 amp, leading to curves with a rather different appearance from those in the textbooks. The curves are shown in Figs. 11(*a*) and 11(*b*), and may be compared with those given in Reference 2, p. 368.

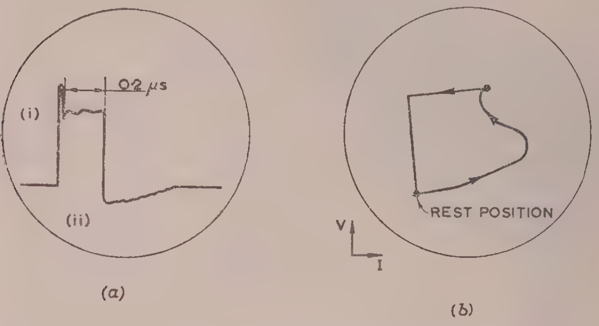


Fig. 11.—Current/time and voltage/current curves.

(*a*) Current pulse. Note capacitive charging-current peak, (i), and capacitive discharge back into bleeder resistance, (ii).
(*b*) Voltage/current trace.

After full testing in the electromagnet, the magnetrons are fitted with the permanent magnets and final acceptance tests are made, to the specification already given.

(11) ACKNOWLEDGMENTS

The author is indebted to the Admiralty, and to Mr. G. E. Condliffe, E.M.I. Research Laboratories, for permission to publish the paper, and to the Superintendent, Services Electronics Research Laboratory, for permission to describe much of the work carried out there. He is also obliged to many colleagues at the E.M.I. Research Laboratories and S.E.R.L. for their helpful discussions and detailed investigations.

(12) REFERENCES

(1) Boot, H. A. H., and Randall, J. T.: "The Cavity Magnetron," *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 928.
(2) Collins, G. B. (Ed.): "Microwave Magnetrons," M.I.T. Radiation Laboratory Series (McGraw-Hill).
(3) Wootton, D. J., and Pearce, A. F.: "A Reflex Klystron for the 8–9 mm Band" (see page 104).
(4) British Patent No. 667575.

(13) APPENDICES

(13.1) Calculation of Cold Pulling-Figure

(i) *First Stage.*

The variable-phase 1·5 : 1 v.s.w.r. is produced by a puller of impedance $Z_p = Z_3/2$ in a matched line of characteristic impedance Z_3 (Fig. 9). The impedance Z_D at the puller is transformed by the usual line equations to Z_C at the transformer and through the two quarter-wave sections to Z_A at the cavity mouth. Inverting to an admittance, we find

$$\begin{aligned} Y_A &= G_A + jB_A \\ &= \frac{Z_2^2}{Z_1^2 Z_3}\,\frac{Z_3(Z_3+Z_p)}{(Z_3+Z_p)^2\cos^2\beta_3 l_3 + Z_3^2\sin^2\beta_3 l_3} \\ &\quad + j\frac{Z_2^2}{Z_1^2 Z_3}\,\frac{[(Z_3+Z_p)^2 - Z_3^2]\sin 2\beta_3 l_3}{2[(Z_3+Z_p)^2\cos^2\beta_3 l_3 + Z_3^2\sin 2\beta_3 l_3]} \quad . \quad (9) \end{aligned}$$

where β_3 is the phase-change coefficient in Z_3. Hence

$$\frac{\partial B_A}{\partial l_3} = \frac{Z_2^2 Z_p(2Z_3+Z_p)}{2Z_1^2 Z_3} \times \frac{2\beta_3(Z_3+Z_p)^2\cos^2\beta_3 l_3 - 2\beta_3 Z_3^2\sin^2\beta_3 l_3}{[(Z_3+Z_p)^2\cos^2\beta_3 l_3 + Z_3^2\sin^2\beta_3 l_3]^2} \quad . \quad (10)$$

neglecting long line effect (i.e. assuming $l_3\delta\beta_3 \ll \beta_3\delta l_3$). This vanishes at the extreme frequency excursions, giving

$$(Z_3+Z_p)^2\cos^2\beta_3 l_3 - Z_3^2\sin^2\beta_3 l_3 = 0 \quad . \quad . \quad (11)$$

Solving for l_3 and inserting in the imaginary part of eqn. (9), the extreme values of B_A are therefore

$$B_A = \pm\frac{Z_2^2 Z_p(2Z_3+Z_p)}{2Z_1^2 Z_3^2(Z_3+Z_p)} = \pm\frac{5}{12}\frac{Z_2^2}{Z_1^2 Z_3} \quad . \quad . \quad (12)$$

The admittance of N resonators of impedance Z_1, a small fraction σ off resonance, is

$$\begin{aligned} NY_r &= j(N/Z_1)\cot\beta_1(1-\sigma)l_1 \\ &\simeq j(N/Z_1)\sigma\pi/2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13) \end{aligned}$$

The comparative agreement of hot and cold measurements justifies the neglect of effects of the electron stream, so that the admittance changes due to load and to cavity detuning must balance. Therefore

$$j(N/Z_1)\sigma\pi/2 = \pm j(5/12)Z_2^2/Z_1^2 Z_3$$

Inserting values for the VX5027, we have

$$\sigma = \pm 0{\cdot}000837 \quad . \quad . \quad . \quad . \quad . \quad (15)$$

The pulling figure to a first approximation is therefore

$$f_p = 2\sigma f_0 = 60\ \text{Mc/s} \quad . \quad . \quad . \quad . \quad . \quad (16)$$

(ii) *Second Stage.*

The admittance formulae given on pp. 62 and 63 o Reference 2 may now be used to correct for the true resonato shapes. The admittance Y_A at the mouth of the output resonato is

$$Y_A = j\left(\frac{\epsilon_0}{\mu_0}\right)^{\frac{1}{2}}\frac{l_a}{\psi a} \times \frac{J_0(\beta a) - \dfrac{J_1(\beta b_0)Y_B - j(\epsilon_0/\mu_0)^{\frac{1}{2}}(l_a/\psi b_0)J_0(\beta b_0)}{N_1(\beta b_0)Y_B - j(\epsilon_0/\mu_0)^{\frac{1}{2}}(l_a/\psi b_0)N_0(\beta b_0)}N_0(\beta a}{J_1(\beta a) - \dfrac{J_1(\beta b_0)Y_B - j(\epsilon_0/\mu_0)^{\frac{1}{2}}(l_a/\psi b_0)J_0(\beta b_0)}{N_1(\beta b_0)Y_B - j(\epsilon_0/\mu_0)^{\frac{1}{2}}(l_a/\psi b_0)N_0(\beta b_0)}N_1(\beta a} \quad . \quad . \quad . \quad . \quad (17$$

in which Y_B is the admittance at B (Fig. 7 or 9) outwards and the other dimensions are as shown in Fig. 7. It is assume that the maximum frequency shifts of the magnetron will sti occur for positions of the puller given by eqn. (11); the corre sponding values of Y_B are found to be

$$Y_B = \frac{Z_3}{Z_2^2}\left(\frac{78}{97} \pm j\frac{30}{97}\right) \quad . \quad . \quad . \quad . \quad (1$$

The odd figure 97 arises as $2^4 + 3^4$, and 78 as $2 \times 3(2^2 + 3^2$

Hence

$$Y_A = 0{\cdot}000915 + j0{\cdot}00278 \text{ mho}$$
$$\text{or } 0{\cdot}000871 + j0{\cdot}003465 \text{ mho}$$

t the extreme frequency excursions. The change of susceptance s

$$\delta B_A = 0{\cdot}000685 \text{ mho} \quad . \quad . \quad . \quad . \quad . \quad (19)$$

The admittance of a resonator is given by

$$Y_r = j\left(\frac{\epsilon_0}{\mu_0}\right)^{\frac{1}{2}} \frac{l_a}{\psi a} \frac{\mathrm{J}_0(\beta a)\mathrm{N}_1(\beta b) - \mathrm{J}_1(\beta b)\mathrm{N}_0(\beta a)}{\mathrm{J}_1(\beta a)\mathrm{N}_1(\beta b) - \mathrm{J}_1(\beta b)\mathrm{N}_1(\beta a)} \quad . \quad (20)$$

This is, of course, the limit of eqn. (17) when $b_0 = b$ and $Y_B \rightarrow \infty$. The variation for fractional detuning is not found by differeniating this expression (the result is unmanageable), but simply y evaluating Y_{rl} and Y_{rs} for the long and short resonators at wo values of λ $(2\pi/\beta)$.

This gives $\dfrac{N}{2}\dfrac{\delta B_{rl}}{\delta\lambda} + \dfrac{N}{2}\dfrac{\delta B_{rs}}{\delta\lambda} = 0{\cdot}0575 \text{ mho/mm} \quad . \quad . \quad (21)$

Hence the wavelength change due to δB_A is 0·0119 mm, and he pulling figure (second approximation) is

$$f_p = 48 \text{ Mc/s}$$

ii) *Third Stage.*

The H transformer has the same cut-off wavelength and the ame gap height as the rectangular slot; the characteristic mpedance Z_{2H} is found to be 17% higher (see Reference 2, . 200), and since the pulling figure varies as Z_2^2, we have a hird approximation

$$f_{pH} = 1{\cdot}17^2 f_p = 66 \text{ Mc/s}$$

(13.2) Effects of Error in Transformer Length

Suppose the transformer is not exactly a quarter-wavelength ong, so that

$$l_2 = (\lambda_2/4)(1 - \kappa) \qquad (\kappa \ll 1)$$

Then $$Y_B \simeq \frac{Z_C}{Z_2^2} + j\frac{\kappa\pi(Z_C^2 - Z_2^2)}{2Z_2^2} \quad . \quad . \quad . \quad . \quad (24)$$

Writing eqn. (17) in the form

$$Y_A = j\frac{pY_B + jq}{rY_B + js} \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

inserting eqn. (24) for Y_B, selecting the imaginary part and dropping small terms $(s^2Z_2^2, qsZ_2^2)$, we have

$$B_A = \frac{\begin{aligned} &pr|Z_c|^2 + (ps + qr)X_cZ_2^2 \\ &\qquad + \kappa\pi[-pr(X_c/Z_2)(R_c^2 + X_c^2 + Z_2^2) \\ &\qquad\qquad + \tfrac{1}{2}Z_2(ps + qr)(R_c^2 - X_c^2 - Z_2^2)]\end{aligned}}{\begin{aligned} &r^2|Z_c|^2 + 2rsX_cZ_2^2 \\ &\qquad + \kappa\pi[-r^2(X_c/Z_2)(R_c^2 + X_c^2 + Z_2^2) \\ &\qquad\qquad + Z_2rs(R_c^2 - X_c^2 - Z_2^2)]\end{aligned}} \quad . \quad . \quad . \quad . \quad (26)$$

where $R_c + jX_c = Z_c$.

In eqn. (26) the first terms of numerator and denominator are larger than all the others, and their ratio p/r measures the fixed frequency-shift caused by even a matched guide $(X_C = 0)$ when the output cavity is not $\lambda/4$ long. On taking the smaller second terms into account we have the cold pulling figure as already found.

The next terms are $-\kappa\pi pr(X_C/Z_2)(R_C^2 + X_C^2 + Z_2^2)$ and $-\kappa\pi r^2(X_C/Z_2)(R_C^2 + X_C^2 + Z_2^2)$, and their ratio p/r is the same as that of the first terms. They therefore do not alter the value of B_A except to a small degree through the presence of the second and fourth terms. The result is a negligible change in f_p.

This does not apply to the last terms; these contain X_c (the reactance at C due to the puller) only as X_c^2. The frequency shift represented by these terms is not negligible, and is in the same direction for all phases of the puller. Thus a simple Rieke diagram of the type shown in Fig. 10(*a*) is distorted to the type shown in Fig. 10(*b*); the direction of curvature is, of course, determined by the sign of κ. Diagrams similar to Fig. 10(*b*) have been observed in practice. The calculated mean frequency-shift at a v.s.w.r. of 1·5 is 600κ Mc/s.

621.385.1.029.6: 621.373.423

Monograph No. 14
Aug. 1

A REFLEX KLYSTRON OSCILLATOR FOR THE 8–9 MM BAND

By D. J. WOOTTON, B.Sc., Associate Member, and A. F. PEARCE, Ph.D., F.Inst.P.

(*The paper was first received* 14*th January, and in revised form* 12*th April*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in August*, 195

SUMMARY

The paper describes a reflex klystron oscillator, tunable over the wavelength range of 8–9 mm, which is suitable both for use in a superheterodyne receiver and as a source for laboratory measurements.

The cavity operates in its fundamental mode and tunes smoothly over the wavelength band. It is mounted in a metal envelope having a glass window through which the output power passes.

At an input of 2 000 volts 10 mA, the output power of an average valve is about 30–50 mW, and the electronic tuning range to half-power is 60 Mc/s. Powers up to 160 mW have been obtained.

The factors affecting the performance of the valve are discussed, and consideration is given to the noise power generated by the oscillator when used in a superheterodyne receiver.

LIST OF PRINCIPAL SYMBOLS

The M.K.S. system of units is used throughout.

β = Beam coupling factor.
R = Shunt resistance of the unloaded resonator.
I = Effective resonator current.
N = Overall noise factor of superheterodyne receiver.
L = Conversion loss of crystal mixer.
t = Noise temperature ratio of crystal (n.t.r.).
t' = Excess n.t.r. due to local-oscillator noise.
$N_{i.f.}$ = Noise factor of intermediate-frequency amplifier.
P_n = Local oscillator noise power contained in the two noise sidebands.
T = Absolute temperature.
B = Bandwidth of receiver.
P = Power output from local oscillator.
f_0 = Natural resonant frequency of the oscillator.
ω = Angular frequency.
F' = A bunching factor which depends upon the potential distribution in the reflector space.
s = Total drift length.
V_0 = Beam potential at the resonator-gap centre.
Q_L = Loaded Q-factor.
ϕ = Difference of phase between the arrival of the electron bunches at the resonator and the maximum electric field across the gap.
C_0 = Equivalent capacitance at the resonator gap.

(1) INTRODUCTION

Since the 1939–45 War attention has been given to the development for radar purposes of still shorter wavelengths than those formerly used, with a view to obtaining, amongst other advantages, aerials of smaller size for a given resolution. One such band selected for exploitation was in the wavelength range of 8–9 mm, where there exists a region of minimum atmospheric absorption. A tunable low-power oscillator for this band, suitable for use in superheterodyne receivers and also for bench measurements, was therefore needed.

The first design to operate in this band was of disc-seal construction, the cavity oscillating in a harmonic mode, and was developed by workers at the Clarendon Laboratory, Oxfo
from an earlier valve for the 1·25 cm band. Samples of t
valve were made both at the Radar Research Establishme
(formerly the Telecommunications Research Establishment) a
at the Services Electronics Research Laboratory, but by late
1947, difficulties had been encountered both in manufacture a
performance (these were later much reduded) and there appear
to be a need for an alternative design. The paper describes t
outcome of the development started at that time, the final va
being known as type VX5023. A magnetron for this band h
also been developed and is described in a companion paper.[1]

Within about six months the basic design had been evolved
means of a series of experiments in a demountable system; the
after attention was turned to making sealed-off valves inc
porating a tuning mechanism, and the development was virtua
complete by 1950.

It was decided at the outset to use a high-voltage input
about 2 kV; this decision was governed mainly by practical co
siderations. The value chosen was thought to be the minimu
that would give the performance required, using a resona
having gridless apertures. The grids that would have be
necessary in a low-voltage valve for this frequency would ha
been very difficult to make owing to the minute dimensio
involved, and might well have been capable of dissipating onl
very limited amount of power, thus limiting the input power a
efficiency. Another advantage of the high-voltage valve, wh
has been found to be of importance in this band during the l
few years, is that local-oscillator noise in a superheterody
receiver is less.

With regard to the cavity, the choice lay between fund
mental- and harmonic-mode operation. The harmonic type h
the advantage of being bigger and therefore easier to tune a
to handle generally. The fundamental type, however, has t
advantages: first, the copper losses are less than for any oth
cavity and the efficiency is therefore greater, and secondly,
frequency may be varied smoothly over a 10% band with
sudden variation of output, whereas this is not always possi
with certain types of harmonic cavity.

The fundamental cavity turned out to be of about 4 r
diameter and 1 mm depth, the apertures being of about 0·5 r
diameter. These dimensions are very small, but experim
showed that, by using a thin diaphragm for one wall of the cav
the capacitance could be varied sufficiently to tune the freque
over about 10%. A cavity of the fundamental type was theref
chosen. More recent developments of harmonic cavities sh
that the disadvantages are not necessarily as great as was belie
in 1947, and such cavities can therefore be useful in oscillator
this type.

On considering possible methods of construction, the prin
necessity was to ensure great accuracy of assembly, particul
with regard to axial alignment of the gun, resonator and refle
It was therefore decided to machine the resonator from a co
block, and to locate the gun and reflector with respect to
block. The copper block was mounted inside a thin
envelope, the oscillatory power being coupled to an ext
waveguide through a glass window of the type which had

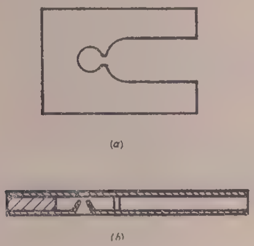
Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Wootton and Dr. Pearce are at the E.M.I. Research Laboratories Ltd.

ed on magnetrons. The resonator was placed in good thermal ntact with the metal envelope, thus ensuring adequate cooling id minimizing frequency drift.

(2) THE BASIC ELECTRICAL DESIGN

The method of arriving at the final design was to determine many of the relevant dimensions as possible by a combination calculation and previous experimental knowledge, and to find e remainder by empirical means.

Several difficulties stand in the way of designing more cometely by calculation, but the principal one is concerned with ectron optics. At the high current densities required in valves the type under consideration, the electron paths are determined ore by mutual repulsion between electrons than by the laws of ometrical electron optics. The reflex klystron is complicated the fact that the electron beam reverses direction near the flector, and under these conditions the electron trajectories nnot be calculated with any accuracy.


(a)

(b)

Fig. 1.—The basic resonator design.

It is important to keep the resonator apertures as small as ssible, consistent with passing substantially the whole of the ectron beam, since the resonator losses are otherwise too big hunt resistance R too low) and the coupling factor (β) too low. e size chosen will of course depend upon the magnitude of the ectron current which it is proposed to use (assuming that the input voltage has already been fixed). The current which would give the required output power was calculated by the method given in Section 5, taking the best estimates possible for quantities such as the shunt resistance R of the resonator. It appeared that a current in the region of 8–12 mA at 2000 volts would be adequate, and for this current a resonator aperture of 0·020 in diameter was chosen.

It is advantageous in a reflex oscillator to have the resonator aperture on the reflector side slightly larger than that on the gun side. By this means a number of marginal electrons that would otherwise be lost are allowed to enter the resonator after reflection and to give up energy, which more than compensates for the lower coupling factor (β). For the larger aperture the diameter chosen was 0·025 in.

The best value for the separation of the two apertures—the "gap"—is given by the condition that $\beta^2 R$ should be a maximum.[2] For small gaps β is high and R low, whilst for large gaps the opposite obtains. In between these extremes a broad optimum exists. β is readily calculable,[3] but, as is shown later, R can be determined only in an approximate manner. Nevertheless the optimum $\beta^2 R$, and hence the optimum gap, may be found by such calculations sufficiently accurately to enable experimental work to start.

The remaining resonator dimensions, namely the diameter and the depth, are easily determined, at least approximately, from existing data. Slight changes may be necessary to get the exact frequency required when the resonator is made up and tested. Some choice of the ratio of diameter to depth is available, and in this case a fairly large ratio was chosen so as to give as large a diameter as possible to the flexible diaphragm.

The first experiments were carried out with a continuously-pumped demountable valve. This was the simplest way to establish a suitable design for the electron gun and reflector, to check the frequency of the cavity, and to make adjustments if necessary. The resonator cavity and output section were machined from a copper plate [Fig. 1(a)], the cavity being completed by two plates in which the apertures had been formed. Good joints were ensured by silver-plating all three parts and

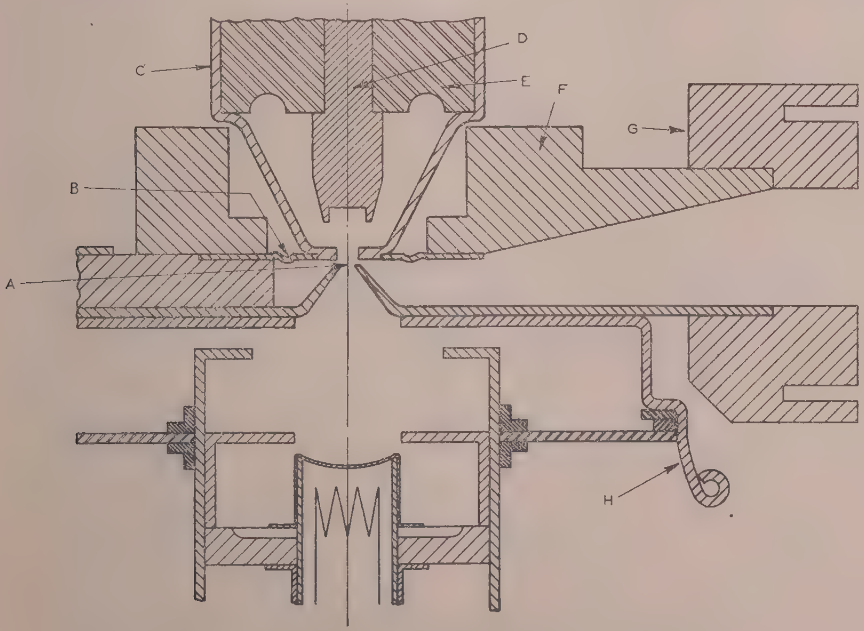


Fig. 2.—Section showing resonator and electrode assembly.

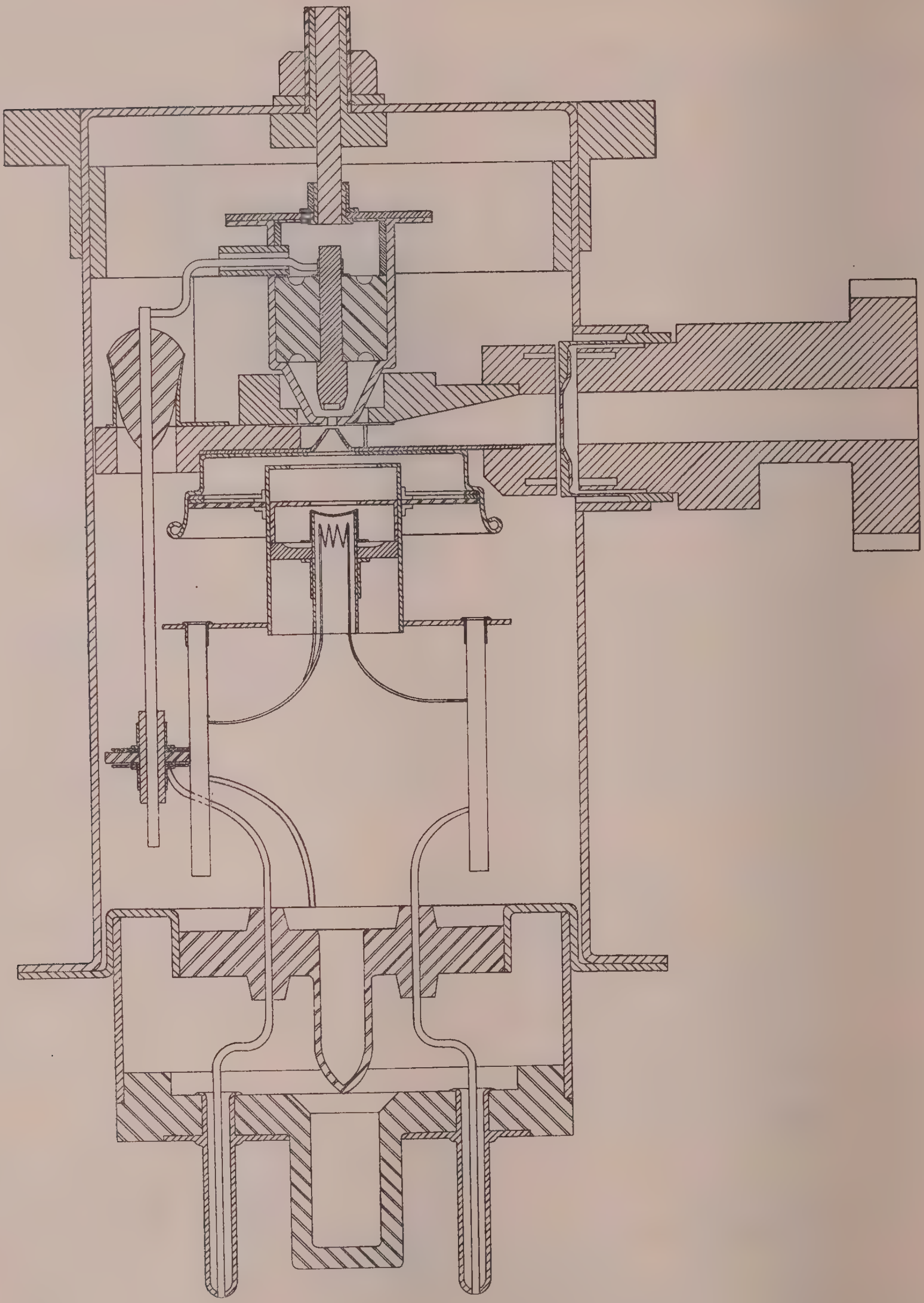


Fig. 3.—Sectional view.

brazing together after assembly in a jig. A sectional view of the assembly is shown in Fig. 1(*b*).

The design of the electron gun was based on an earlier design of an oscillator for the 3 cm band,[4] but was modified for this application. The oxide cathode was of 3 mm diameter and had a concave surface of radius of curvature $\frac{3}{16}$ in. This was surrounded by a focusing electrode, of the shape shown in Fig. 2, which normally operated at a negative potential with respect to the cathode. It was found that the efficiency of transmission through the resonator aperture was about 90%.

The reflector was of the conventional dish shape, and of 0·050 in diameter and 0·025 in depth. Other dimensions were tried in the demountable system before these values were decided.

Whilst these experiments were being carried out it was found that very accurate alignment of the electrodes and apertures was necessary. If the reflector were as little as 0·002 in off the position for maximum power, the output dropped to one-half. The gun position was not so critical.

After a series of demountable-system experiments, a performance fairly close to that expected was obtained, and the decision to start making sealed-off valves was taken. During this stage the technique of assembling the valve and mounting it within the metal envelope in a vacuum-tight manner was studied. Sundry small changes were incorporated; these included slight variation of the output coupling slot and variation of the gap for maximum output until a satisfactory performance was obtained.

(3) CONSTRUCTION

(3.1) The Resonator Assembly

The resonator assembly previously described was adapted for use in the sealed-off valve, in particular to provide means of tuning. Various ways of tuning were considered before it was decided to adopt the method of capacitance variation by flexing one wall of the resonator. The tuning diaphragm B (Fig. 2) was corrugated to improve flexibility and joined to a nickel tube C, which housed the reflector D, and could be rigidly connected by brazing to the top of the valve envelope. This was sufficiently flexible to be capable of being moved from outside by means of the mechanism described later. The reflector was a metallic rod with a cup-shaped portion machined concentrically in the end facing the resonator diaphragm, and was held in place by an accurately ground ceramic insulator E. With this construction the reflector can be located on the axis to within 0·001 in. A taper from the resonator output coupling slot up to the full depth of the standard waveguide was formed in the copper block F; the latter also acted as a stop to prevent the diaphragms from being damaged during tuning.

(3.2) The General Assembly

A sectional view of the valve (less tuner) is shown in Fig. 3. The output window consisted of a disc of glass sealed across an aperture in a cup-shaped pressing made of iron-nickel-cobalt alloy, which was brazed to a tubular metal side-arm on the envelope. On the outside, a short length of waveguide was soldered into the window cup, and was terminated at its far end in a standard coupler. The waveguides facing the window on either side were fitted with ditch chokes—so as to avoid loss of output power. The resonator assembly was attached by means of screws to a metal rim welded to the envelope. By using a jig when joining the rim to the envelope, accurate location of the resonator assembly and choke G (Fig. 2) with respect to the window was ensured. When the dimensions of the window iris, glass thickness and spacings between the window and two ditch chokes were correctly maintained, the voltage standing-wave ratio of the output run was less than 1·2 over the wavelength range of 8–9 mm. A photograph of the valve complete with tuning mechanism is shown in Fig. 4.

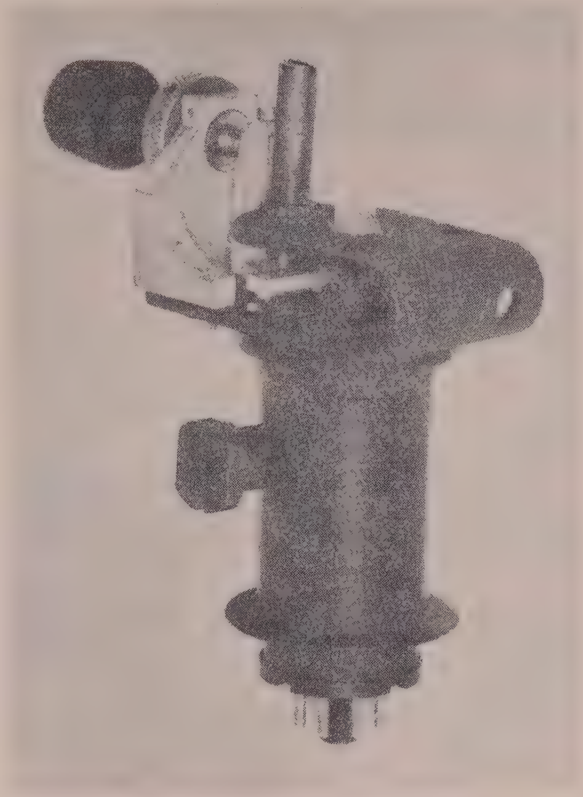


Fig. 4.—Photograph of valve type VX.5023.

The remaining technical problems which arose in the design of the sealed-off valve were concerned with the attachment of the gun and electrical connecting wires to the resonator, and the technique of making the final vacuum seal between the pinch and valve envelope.

The pinch consisted of a seal between a lead-glass disc, through which six dumet leads were sealed, and a nickel-chrome-iron alloy surround having a flat flange to mate with a similar one on the steel envelope. These were joined together by arc welding, and this method was found to be convenient and satisfactory for small numbers of valves.

A solution to the difficulty of attaching the gun and leads to the resonator was found by flexibly mounting the gun on the pinch so that it could be sprung into the nickel dish, shown at H in Fig. 2, the latter being brazed concentrically on to the base of the resonator assembly. The reflector lead from the pinch is at the same time joined by means of a sliding contact to an insulated lead passing through the resonator block. This arrangement is satisfactory since the reflector takes no current.

(3.3) The Mechanical Tuning Mechanism

The natural frequency of the cavity is a very sensitive function of the setting of the resonator gap, a change of 0·001 in giving rise to a frequency shift of 500 Mc/s. It is necessary, therefore, that the external tuning mechanism should provide a large velocity ratio for ease in tuning. This has been obtained by use of a screw-and-lever mechanism in combination with a C spring, the latter being connected to the centre of the top of the valve envelope, which flexes during tuning. The mechanism, which can be seen mounted on top of the valve in Fig. 4, is similar to that originally used on the 1¼cm oscillator (type VX302).

(4) CHARACTERISTICS

Table 1 gives the operating conditions of the valve.

Table 1

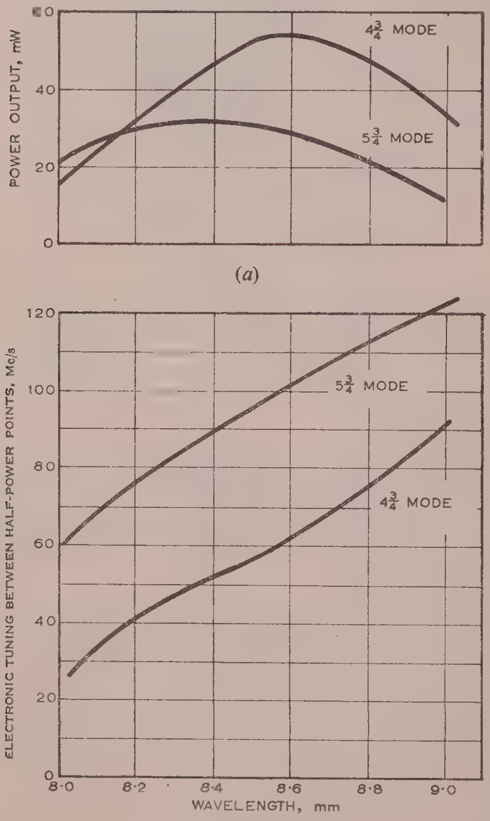
Operating Conditions

Heater voltage	6·3 V
Heater current	0·9 A
Cathode–resonator voltage . .	2·0 kV
Resonator current . . ∴	8 to 12 mA
Screen–cathode voltage range . .	0 to −200 V
Reflector–cathode voltage range	−100 to −500 V

The resonator current is adjusted for maximum output by means of the cathode–screen voltage.

(4.1) Power Output and Electronic Tuning Range

The variation of power output and electronic tuning range with wavelength is shown in Figs. 5(*a*) and 5(*b*) for two different electronic modes. The curves refer to a valve where the output slot coupling the cavity to the waveguide is 0·060 in wide. The performance varies somewhat over the wavelength band, but at the centre a power output of 50 mW and an electronic tuning range of 60Mc/s is obtained for the 4¾ mode.



(*a*)

(*b*)

Fig. 5.—Characteristics.

A series of demountable-system experiments was performed in order to determine the output-slot width for optimum power and electronic tuning range. The results are shown in Fig. 6, where curves (*a*), (*b*) and (*c*) show the variation of loaded Q, electronic tuning range and power output, respectively, with slot width. It will be seen that the slot size of 0·060 in is undercoupled, maximum power and electronic tuning range occurring at a slot width of between 0·080 in and 0·090 in; the loaded Q is then in the neighbourhood of 400. Another consideration that must be taken into account in determining the best slot size is that of noise output, dealt with in more detail in Section 4.2. It is known that noise output increases with slot width (decreasing Q_L), and for some applications an oscillator having the full slot width might generate too much noise. A slot width of 0·060 in, and more recently 0·070 in, has therefore been used. It may be found possible to increase this to the optimum value when more is known about the cause of the wide fluctuations in noise output from valve to valve mentioned later, and about what level of local oscillator noise can be tolerated in practice in a receiver for the 8–9 mm band.

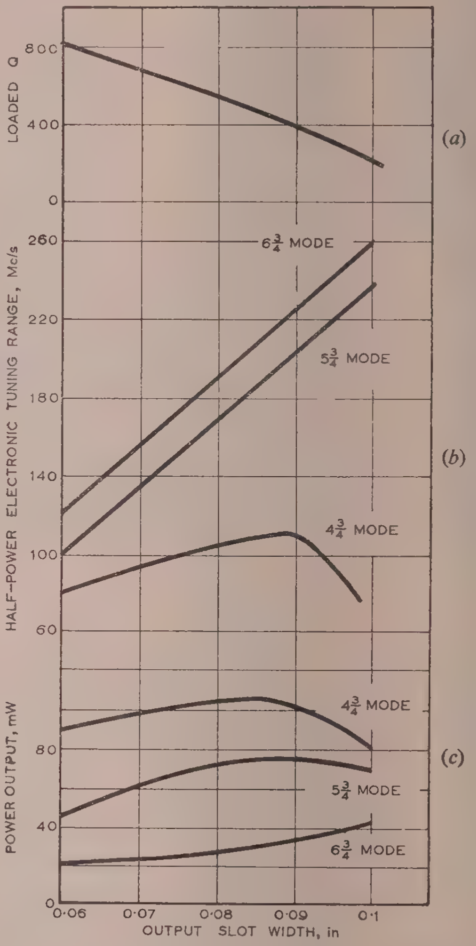


Fig. 6.—Variation of characteristics with cavity loading.

For the measurement of power an enthrakometer bridge[5] was used. Some results on early valves were obtained with a thermistor bridge, but this was replaced after it had been shown[6] that it is subject to an absolute error which becomes appreciable at wavelengths below 3 cm.

Table 2

VARIATION OF RECEIVER NOISE FACTOR WITH EXCESS NOISE TEMPERATURE RATIO

t'	0	0·2	0·3	0·5	1	2	3	4	5	10
N(dB)	13·2	13·5	13·6	13·9	14·5	15·4	16·2	16·9	17·5	19·6

The electronic tuning range was measured using a sine-wave modulation on the reflector electrode, since it was found that unreliable results were obtained if frequencies corresponding to the half-power points were measured by manual control of the reflector voltage. This was due to a slow heating effect caused by a small change in resonator current as the reflector voltage was varied, and resulted in a frequency change in addition to that due to normal electronic tuning.

(4.2) Local-Oscillator Noise

The local oscillator in a superheterodyne receiver generates a certain amount of noise. This is of little importance at wavelengths of 3 cm and above but can be significant in the 8–9 mm band. The local-oscillator contribution to receiver noise is conveniently expressed as an increase t' in the noise temperature ratio of a crystal mixer. The overall noise factor of a superheterodyne receiver at centimetric frequencies is given by

$$N = L(t + t' + N_{i.f.} - 1) \quad . \quad . \quad . \quad . \quad (1)$$

All these factors have to be reduced as far as possible for good receiver performance.

The envelope of the noise spectrum generated by the valve is shown in Fig. 7, the two noise sidebands which contribute to receiver noise being spaced at the intermediate frequency from that of the local oscillator. Clearly the magnitude of the noise power depends on the intermediate frequency and the amplifier bandwidth, as well as on the local-oscillator noise spectrum, the latter being dependent on the loaded Q of the resonator.

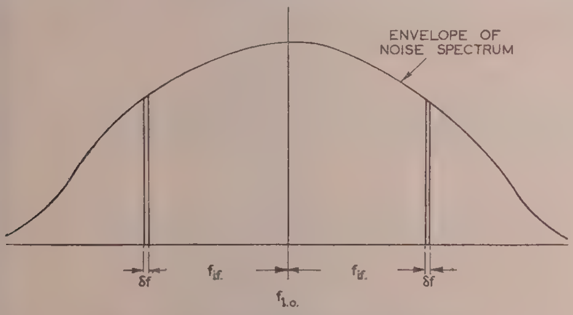


Fig. 7.—The local-oscillator noise spectrum.

From elementary considerations it may be shown that

$$t' = \frac{a}{kT}\frac{P_n}{BP} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where a is a constant and k is Boltzmann's constant.

Since P_n is proportional to the bandwidth, t' is a measure of the noise to signal-power output from the local oscillator per unit bandwidth. It will increase as the loaded Q of the cavity is decreased and with decrease of intermediate frequency.

It is instructive to determine how the noise factor of a receiver for use in the 8–9 mm band varies with t' if reasonable values of the other receiver parameters are inserted in eqn. (1). Some values are given in Table 2 for a single crystal mixer making the following assumptions: $L = 8 \cdot 5$ dB; $t = 2 \cdot 0$; $N_{i.f.} = 3$ dB (for 45 Mc/s amplifier).

Table 2 shows that the deterioration in receiver performance using a simple mixer would begin to become excessive if t' were allowed to exceed 0·5. The effect of local-oscillator noise can, however, be reduced considerably by the use of a balanced mixer, and if the crystals are perfectly matched, local-oscillator noise is completely cancelled. In practice, however, some mismatch occurs, and so it remains important to maintain as low a value of local-oscillator noise as possible. Just what level can be accepted when using a balanced mixer has not yet been determined, but it is thought that a maximum excess noise temperature ratio of about 10 is permissible. If the match between the crystals is poor, the noise-suppression ratio, which is the amount by which the local-oscillator noise is reduced by use of the balanced mixer, might fall to 20. In these circumstances the deterioration in receiver performance, due to a local oscillator where $t' = 10$, is about 0·7 dB. Normally the deterioration would be considerably less.

No apparatus was available to measure the excess noise temperature ratio at the time of the main development of the valve, but subsequently measurements were made by C. R. Ditchfield of R.R.E., Malvern. Average results for different reflector modes are given in Table 3 for the valve with the 0·060 in output coupling slot.

Table 3

MEASUREMENTS OF EXCESS NOISE TEMPERATURE RATIO

Reflector mode	4$\frac{3}{4}$	5$\frac{3}{4}$	6$\frac{3}{4}$
t' at mode centre	1·7	4·1	10
t' at the low-frequency half-power point	4·3	10·4	>20
t' at the high-frequency half-power point	5·3	16·8	>20

The increase in t' with increasing electronic mode number is due to the fact that the noise power remains sensibly constant with mode, whereas the power output decreases with increasing mode number over the range considered.

Kuper and Knipp (Reference 7, Chapter 17) have discussed the general noise behaviour of the reflex klystron using the shot effect in the injected current as a basis for noise generation. They showed that the asymmetry of the noise with regard to the half-power tuning points may be explained by the change of relative phase between the injected and reflected noise currents as the reflector voltage is varied. This effect is demonstrated in Table 3.

While the general noise behaviour of the VX5023 can be explained in terms of known theory, wide fluctuations in noise output have been observed. For instance, in a batch of valves where the cavity loadings were all approximately the same, it was found that individual values of t' ranged from 1·5 to over 10 for the centre of the 5$\frac{3}{4}$ electronic modes, the mean value being 4·1. The precise explanation for this is not yet known, but experiments now in progress show that the variation is due partly to displacement of the electrodes.

(5) THEORETICAL

The efficiency and electronic tuning range of a reflex klystron may be calculated in a straightforward manner using a small-signal approximation, if the values of several parameters are known. Some of these, however, are known only approximately; moreover, the theory used ignores some effects that may be appreciable, such as the effect of space charge on bunching of the beam. Consequently the numerical values obtained are somewhat approximate; nevertheless the calculations were of considerable assistance in deciding certain features of the design.

The calculations were based on the theory of Barford and Manifold.[2]

(5.1) Efficiency

The efficiency is given in terms of two parameters C and k, where

$$C = 1{\cdot}68 \times 10^{-6} \times \omega F' s V_0^{-1/2} \quad . \quad . \quad . \quad . \quad (3)$$

and

$$k = \beta^2 RIC V_0^{-1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The efficiency is inversely proportional to C, and increases with k from zero at $k = 1$, steeply at first, but flattening off for large values of k (see Reference 2, p. 305). For high efficiency C must be small and k large.

In these equations F', s and V_0 may be obtained from a field plot using an electrolytic trough, although the value of F' found in this way may be to some extent in error because of the neglect of space charge. β can be calculated with fair accuracy,[3] the value of 0·6 being obtained for a resonator gap spacing of 0·006 in. The effective current is another factor which is difficult to assess accurately. It has been taken here as 80% of the total current, i.e. 8 mA.

The shunt resistance of the resonator, which is the resultant of contributions from copper losses and beam damping, is difficult to assess with any accuracy. The part due to beam damping can be calculated, and may be neglected since it is greater than 1 megohm. Published data (Reference 7, p. 78) for a resonator of approximately the shape in use in this oscillator but without apertures gives a shunt resistance due to copper losses of 50 000 ohms. The presence of apertures would perhaps raise this value, but on the other hand the surface conductivity of copper rarely equals the low-frequency value, so it is likely that a somewhat lower value than 50 000 ohms would be obtained in practice.

If, for the moment, the shunt resistance is regarded as a variable, the efficiency calculated from eqns. (3) and (4) can be plotted as a function of R. This has been done in Fig. 8 for the 4¾ and 5¾ modes. Then, if the theoretical efficiency is to agree with the maximum experimental value obtained (0·8% for the 4¾ mode), R must be in the region of 30 000 ohms, which is in reasonable agreement with the estimate made above. Using this value of shunt resistance a value of theoretical efficiency for the 5¾ mode is obtained which is greater than the measured one. This appears to be a general characteristic of the reflex klystron, and has been explained[8] in terms of the increased drift time for the higher mode. This is liable to cause an increased loss of electrons after reflection, and to increase any phase difference that may exist between the axial and the marginal electrons.

(5.2) Electronic Tuning Range

It is well known that the frequency of oscillation f, when the reflector voltage is slightly different from that giving maximum power, is given by

$$f - f_0 = -\frac{f_0}{2Q_L} \tan \phi \quad . \quad . \quad . \quad . \quad (5)$$

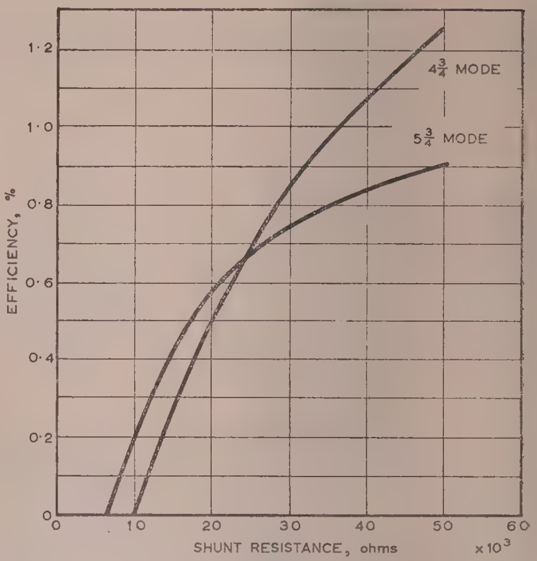


Fig. 8.—Valve efficiency as a function of shunt resistance.

It has been shown[8] that

$$Q_L = \frac{C_0}{1{\cdot}68 \times 10^{-6} \times \beta^2 IF' s V_0^{-3/2} \dfrac{2J_1(x)}{x}} \quad . \quad . \quad (6)$$

$\dfrac{2J_1(x)}{x}$ is a function depending solely upon k, and can therefore be computed from the known value of k. C_0 may be calculated from

$$Q_0 = \omega_0 C_0 R$$

Hence Q_L may be calculated from eqn. (6), and finally the electronic tuning range from eqn. (5), since $\tan \phi$ is a known function of k.[8]

Table 4

Electronic Tuning-Range Calculations

Mode	Half-power electronic tuning range		
	$R = 20\,000$	$R = 30\,000$	$R = 50\,000$
	Mc/s	Mc/s	Mc/s
3¾	43	53	54
4¾	72	77	81
5¾	125	130	138

The results given in Table 4 apply to optimum loading conditions and should therefore be compared with experimental values obtained for a slot size of about 0·090 in. It will be seen that the calculated electronic tuning range is not very sensitive to what value is chosen for the shunt resistance over the range 20 000 to 50 000 ohms. Experimental values obtained, namely 110 Mc/s and 200 Mc/s, are rather greater than the computed ones for the 4¾ and 5¾ modes. It is likely that the effect of space charge on some of the parameters involved, such as the bunching factor F', contributes to the discrepancies.

Considerable variation in electronic tuning range is found between valves. Average values are about 60 Mc/s for the 4¾ mode and about 80 Mc/s for the 5¾ mode for an output-slot width of 0·06 in. These values are smaller than the calculated ones because the load coupling is less than optimum.

(6) ACKNOWLEDGMENTS

The authors wish to thank the Admiralty, under whose auspices the oscillator was developed, and Mr. G. E. Condliffe, Managing Director of E.M.I. Research Laboratories Ltd., for permission to publish the paper. Thanks are also due to Dr. L. F. Broadway for helpful criticism, and to the authors' assistants.

(7) REFERENCES

(1) Vaughan, J. R. M.: "A Millimetre-Wave Magnetron" (see page 95).

(2) Barford, N. C., and Manifold, M. B.: "Elementary Theory of Velocity-Modulation Oscillator," *Journal I.E.E.*, 1947, **94,** Part III, p. 302.

(3) Fremlin, J. H., Gent, A. W., Petrie, D. P. R., Wallis, P. J., and Tomlin, S. G.: "Principles of Velocity Modulation," *ibid.*, 1946, **93**, Part IIIA, p. 890.

(4) Pearce, A. F.: "A Velocity-Modulation Reflection Oscillator for Wavelengths of about 3·2 cm," *ibid.*, 1948, **95**, Part III, p. 415.

(5) Collard, J.: "The Enthrakometer: An Instrument for Measurement of Power in Rectangular Waveguide," *ibid.*, 1946, **93,** Part IIIA, p. 1399.

(6) Collard, J., Nicoll, G. R., and Lines, A. W.: "Discrepancies in the Measurement of Microwave Power at Wavelengths below 3 cm," *Proceedings of the Physical Society*, B, 1950, **63**, p. 215.

(7) Hamilton, D. R., Knipp, J. K., and Kuper, J. B. H.: "Klystrons and Microwave Triodes" (McGraw-Hill Book Co., 1948).

(8) Pearce, A. F., and Mayo, B. J.: "The Design of a Reflex-Klystron Oscillator for Frequency Modulation at Centimetre Wavelengths," *Proceedings I.E.E.*, Paper No. 1301 R, April, 1952 (**99**, Part IIIA, p. 445).

621.3.013.78

Monograph No. 144
Aug. 1955

THE MAGNETIC SCREENING EFFECT OF IRON TUBES

By P. HAMMOND, M.A., Associate Member.

(*The paper was first received* 15*th December*, 1954, *and in revised form* 15*th April*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in August*, 1955.)

SUMMARY

The magnetic screening effect of cylindrical iron tubes of different thicknesses and permeabilities is discussed and analysed, particular attention being paid to the effect of saturation. The problem is approached from a consideration of the induced pole strength on the surface of the iron. The distribution of pole strength is expressed by means of a Fourier series, and it is shown that such a distribution will, in general, cause a magnetic field which varies from place to place in the screened region. Values of the theoretical screening ratio are calculated and good agreement is observed with values obtained experimentally.

(1) INTRODUCTION

It is a well-known experimental fact that the magnetic field in any space can be considerably reduced by surrounding that space by an iron container, so long as the currents and magnets causing the field are external to the container. This property of iron is often referred to as a screening effect, and analogies to the screening effect of an opaque body on light, or of a solid body on a current of air, spring readily to mind. Such analogies may be helpful up to a point, but they do not give a correct picture of the mechanism underlying the magnetic screening effect of iron. It is the aim of the paper to examine the problem from the viewpoint of magnetic theory, which starts from the concept of the inverse square law of force between point magnetic poles.

In using this approach, it is not thought that this is the only way of looking at the problem. In principle, magnetic forces can be evaluated by using the theory of relativity applied to moving charges, but the details of this method do not appear to have been worked out sufficiently to deal with problems of this nature. The classical theory of magnetism, on the other hand, is convenient for mathematical treatment and familiar to engineers. These are the reasons for using it in the paper.

It was stated above that magnetic theory can be built up from the fundamental law that the force between poles varies as the inverse square of the distance. Here it seems desirable to refer to the use in the paper of the inverse square law. Some writers use this law in such a way as to make it dependent on the medium in which it acts, while others state clearly that the law is independent of the medium. The latter view is taken here, and any apparent change due to a different medium is explained by the effect of induced polarity and in particular by the polarity induced on the boundary surface between different media. Attention is withdrawn from the medium and concentrated on the boundary. In the particular case of iron in a magnetic field the change of magnetic force inside a cavity in the iron is thus attributed to the induced surface polarity on the iron. To calculate the screening effect of a particular iron container it is therefore necessary to calculate the pole strength on the surface of the iron.

In all such calculations it is useful to employ the concept of permeability which, in our view, is a convenient device to arrive at a statistical summation of the effects of many induced poles. But in using this device one is faced at once by the complication that the permeability of iron is by no means constant. If the permeability varies, it is likely that the magnetic force H in the iron will also vary. Any such variation of magnetic force must, in the view adopted here, be attributed to a suitable distribution of pole strength on the surface of the iron. Consider the special case of a long cylindrical iron tube placed with its axis transversely to a uniform magnetic field (Fig. 1).

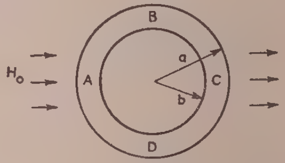


Fig. 1.—Tube in uniform magnetic field H_0.

It is well known that the flux density in the iron will vary around the tube, being zero at A and C and having its maximum value at B and D. Hence the permeability and the magnetic force will also vary, and so will the surface pole strength. The difficulties thus created by the varying permeability of the iron are attacked in the paper by analysing the distribution of surface pole strength by means of a Fourier series. The analysis is found to be competent to predict the screening effect of tubes of any thickness made of iron of widely differing magnetic quality. The theoretical treatment is checked by measurements on a mild-steel tube and a Mumetal tube. It is thought that the method of the paper presents a powerful tool for dealing, not only with this special case of a long tube, but also with many other iron problems in which it is desired to analyse the effect of varying permeability, especially where highly saturated iron is used.

(2) THEORETICAL INVESTIGATION

(2.1) Long Cylindrical Tube of Constant Permeability in Transverse Magnetic Field of Strength H_0

Before we attempt to deal with the effects of varying permeability it is desirable to obtain expressions for the induced pole strength on the walls of a tube of constant permeability. Consider first the induced pole strength on the surface of a *solid* circular cylinder of infinite permeability which is placed in a uniform transverse field H_0. Inside the cylinder the field must be zero, and hence the surface pole strength must there produce a uniform field $-H_0$ to remove the effect of the external field, H_0. If the induced surface pole-strength per unit area is denoted by $\mu_0\sigma$, it will give rise to a magnetic force $2\pi\mu_0\sigma$ at points close to the surface of the cylinder. Hence by the conditions of continuity of the radial field at the cylinder surface we have $2\pi\sigma = H_r$, where H_r is the resolved part of H_0 perpendicular to the surface of the cylinder.

H_r can be written $H_0 \cos\theta$, where θ is the angle between the

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Hammond is in the Engineering Department, University of Cambridge.

irection of H_0 and the radius of the cylinder at the point where ne surface pole strength is $\mu_0\sigma$. Hence

$$2\pi\sigma = H_0\cos\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

nd it is seen that the pole strength per unit area will vary as os θ. Since we are dealing with long cylinders, in order to estrict the problem to two dimensions, we can say that the pole trength has a line density varying as $\cos\theta$. It has thus been hown that a distribution of pole strength varying as $\cos\theta$ roduces a uniform field within the cylinder. In the special case f a cylinder of infinite permeability this uniform field is equal nd opposite to the inducing field. But if the permeability is nite there will still be some field in the cylinder. The realization hat a distribution of pole strength varying as $\cos\theta$ always roduces a uniform field within the cylinder is of great importance.

Consider now the tube of Fig. 1. If the line density of pole trength on the outer surface is denoted by $\mu_0\sigma_a$ and that on the nner surface by $\mu_0\sigma_b$, the conditions of continuity are

$$H_a + 2\pi\sigma_a = \mu(H_a - 2\pi\sigma_a) \quad . \quad . \quad . \quad . \quad (2)$$

t radius a

nd

$$H_b + 2\pi\sigma_b = \mu(H_b - 2\pi\sigma_b) \quad . \quad . \quad . \quad . \quad (3)$$

t radius b

here H_a is due to the radial component of the field H_0 and hat of the pole strength σ_b, and where H_b is due to the radial omponent of the field H_0 and that of the pole strength σ_a. Now the radial component of H_0 is $H_0\cos\theta$, so that both σ_a nd σ_b must vary as $\cos\theta$ and can be written as $\sigma_{a1}\cos\theta$ and $_{b1}\cos\theta$.

Hence $$2\pi\sigma_{a1}\cos\theta = \frac{\mu-1}{\mu+1}H_a = yH_a \text{ (say)} \quad . \quad . \quad (4)$$

he field due to $\mu_0\sigma_b$ at the radius a will be

$$2\pi\frac{b^2}{a^2}\sigma_{b1}\cos\theta$$ [1]

Hence $$2\pi\sigma_{a1}\cos\theta = y\left(H_0\cos\theta - 2\pi\frac{b^2}{a^2}\sigma_{b1}\cos\theta\right) \quad . \quad (5)$$

If $\frac{b}{a} = x$, (say)

$$2\pi\sigma_{a1} = y(H_0 - 2\pi x^2\sigma_{b1}) \quad . \quad . \quad . \quad . \quad (6)$$

imilarly $$2\pi\sigma_{b1} = y(H_0 - 2\pi\sigma_{a1}) \quad . \quad . \quad . \quad . \quad (7)$$

herefore $$2\pi\sigma_{a1} = \frac{y(1-yx^2)}{1-y^2x^2}H_0 \quad . \quad . \quad . \quad . \quad (8)$$

nd $$2\pi\sigma_{b1} = \frac{y(1-y)}{1-y^2x^2}H_0 \quad . \quad . \quad . \quad . \quad . \quad (9)$$

s the permeability tends to infinity, y tends to unity

nd

$$2\pi\sigma_{a1} \to H_0$$
$$2\pi\sigma_{b1} \to 0$$

here will then be no field inside the tube, which is the condition f perfect screening.

It should be noted that, so long as the applied field H_0 is uniorm and the permeability of the tube is constant, the surface ole strength will have a line density varying as $\cos\theta$. Hence, s discussed above, the field in the space surrounded by the tube ill be constant in magnitude and direction, and will act in the irection of the applied field. Its magnitude will be

$$_i = H_0 - 2\pi\sigma_{a1} + 2\pi\sigma_{b1}$$
$$= H_0\left[1 - \frac{y(1-yx^2)}{1-y^2x^2} + \frac{y(1-y)}{1-y^2x^2}\right] \quad . \quad (10)$$

Hence $$H_i = H_0\left(\frac{1-y^2}{1-y^2x^2}\right) \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Hence $$\frac{H_0}{H_i} = \frac{1-\left(\frac{\mu-1}{\mu+1}\right)^2\frac{b^2}{a^2}}{1-\left(\frac{\mu-1}{\mu+1}\right)^2} \quad . \quad . \quad . \quad . \quad (12)$$

$$\simeq \frac{(a+b)(\mu t + a + b)}{4a^2} \quad . \quad . \quad . \quad . \quad (13)$$

provided that $$\mu(a+b) \gg t$$

where $$t = a - b$$

Hence $$\frac{H_0}{H_i} \simeq \frac{\mu t}{2a} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

if $$a \simeq b \text{ and } \mu t \gg (a+b)$$

But H_0 is the strength of the magnetic field before the tube is placed in position, and H_i is the magnetic field in the space surrounded by the tube. In this space the field has therefore been reduced from H_0 to H_i, and we can term the ratio H_0/H the screening ratio of such a tube. In the case of the thin tube under consideration the screening ratio is therefore approximately $\mu t/2a$.

Consider now the flux per unit length carried by the wall of the tube. This can be obtained by finding the tangential magnetic force at any radius r within the tube wall, multiplying this force by the permeability of the tube to obtain the tangential flux density B and integrating this over the thickness of the tube wall.

Thus $$B = \mu\mu_0\left[H_0 - 2\pi\sigma_{a1} + 2\pi\sigma_{b1}\left(\frac{b}{r}\right)^2\right]\sin\theta \quad . \quad (15)$$

and $$\phi = \int_b^a Bdr = \mu\mu_0\left[(H_0 - 2\pi\sigma_{a1})t + 2\pi\sigma_{b1}\frac{b}{a}t\right]\sin\theta \quad (16)$$

$$= \mu\mu_0 H_0 t\frac{(1-y)}{(1-yx)}\sin\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

$$= \frac{2\mu\mu_0}{\mu+1}\,\frac{1}{1-\frac{\mu-1}{\mu+1}\frac{b}{a}}H_0 t\sin\theta \quad . \quad . \quad . \quad . \quad . \quad (18)$$

$$= \frac{2\mu_0 a H_0\sin\theta}{1+\frac{a+b}{\mu t}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

$$= 2\mu_0 a H_0\sin\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

if $$\mu t \gg a + b$$

Therefore, as long as the condition of eqn. (20) holds, the walls of the tube will carry twice that flux which would have crossed the diametral plane of a cylinder of radius a, if the tube had not been there.[2]

The average tangential flux density in the wall of a thin tube will be

$$B = \frac{2\mu_0 a H_0\sin\theta}{t} \quad . \quad . \quad . \quad . \quad . \quad (21)$$

(2.2) Application to Iron Tube in Uniform Magnetic Field

There are two objections that have to be met before these results, which apply to a tube of constant permeability, can be extended to deal with an iron tube in which the permeability will be far from constant.

The first objection raises the question of whether the field of the elemental magnets of the iron is equivalent to the field of a distribution of pole strength over the surface of the iron. This question can best be answered by reference to Green's theorem,[3] which states that any distribution of magnetic dipoles can be replaced by a surface distribution of pole strength combined with a volume distribution of pole strength. The surface distribution of pole strength will be proportional to the intensity of magnetization, while the volume distribution is proportional to the divergence of this intensity. Therefore, as long as the permeability is constant there will be no such volume distribution. Thus the analysis of the previous Section is strictly correct. But even if the permeability varies, the volume distribution will be considerably smaller than the surface pole strength, and good agreement with experimental results is to be expected, if the volume pole-strength distribution is neglected in all iron problems. There can never be any violent rate of change of permeability, unless of course there are lumps of slag in the iron. But then there is in effect another free iron surface within the material. Thus a theory of magnetism that looks for the seat of the action on the surface of the material only will be both mathematically convenient and also firmly based on a close insight into the physical mechanism involved.

The second objection that has to be met is based on the fact that the permeability in iron varies so much. It has been said that an analysis based on the assumption of constant permeability is at best only an academic exercise, and at worst it is liable to lead to disastrous errors in engineering calculation. However, consideration of our problem will show that the assumption of constant permeability, if used intelligently, is most helpful in giving general guidance to the solution of the problem. The reason for this is seen most readily by reference to eqn. (4), in which no such assumption has been made. It is clear that the ratio $(\mu - 1)/(\mu + 1)$ cannot be appreciably affected by a change of permeability of even 10 : 1, so long as μ has a value of at least a few hundred.

It is seen that a general principle emerges from our calculations. All magnetic forces or pole strengths that are functions of μ, in which the numerator and denominator contain powers of μ of the same order, will be practically independent of μ, in any problem involving iron. On the other hand, those magnetic quantities, which are functions of μ in which either the numerator or the denominator contains powers of μ to a higher order, will vary with the state of saturation of the iron. It appears that this principle merely states the obvious, but by making use of it in calculations based on the assumption of constant permeability, results which are far from obvious are obtained.

With reference to the particular case of the iron tube in a uniform magnetic field, we notice that the flux and the flux density in the wall of the tube will be practically independent of μ, while the flux density in the screened space surrounded by the tube will be almost inversely proportional to μ and will therefore depend dominantly on the state of saturation of the iron. In this case, B in iron does not depend on μ, while B in air does. This rather surprising result should be compared with that obtained for an iron tube magnetized by a loop of current,[1] for then the B in the iron depends on μ, while B in the air (the leakage flux) is independent of μ.

The fact that B in the tube wall is practically independent of the permeability of the iron, until this is very heavily saturated, is borne out by the experimental investigation [see Figs. 6(*a*), 11 and 15(*c*)]. But it may be helpful to examine the mechanism by which this comes about.

Consider first the magnetic forces inside the wall of a tube of constant permeability. Fig. 2(*a*) shows a shell of pole strength $\mu_0\sigma_{a1} \cos\theta$ in a uniform magnetic field. The magnetic force in the space surrounded by the shell will be $H_0 - 2\pi\sigma_{a1}$ and will be in the same direction as H_0. Fig. 2(*b*) shows a shell of pole strength $-\mu_0\sigma_{b1}\cos\theta$. The magnetic force outside the shell, but close to it, will be of constant strength $2\pi\sigma_{b1}$, but its direction will be as indicated in Fig. 2(*b*). If the field distributions of Figs. 2(*a*) and 2(*b*) are superimposed we arrive at the case of a tube of constant permeability (see Fig. 3), and have thus seen how the combination of the magnetic force of the external field with the magnetic forces of the induced polarities *guides* the flux around the wall of the tube. From the point of view adopted in this paper, this is an instructive example of the manner in which the polarity of the iron adjusts itself so as to render any iron path a good *conductor* of flux.

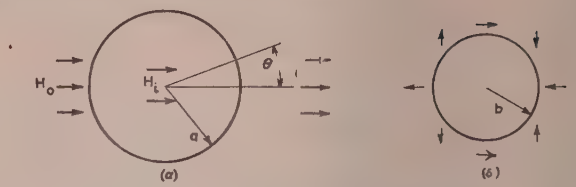


Fig. 2.—Shell of magnetic pole strength.

(*a*) $\mu_0\sigma_{a1}\cos\theta$.
(*b*) $\mu_0\sigma_{b1}\cos\theta$.

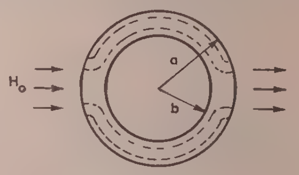


Fig. 3.—Flux guided around walls of tube.

Consider now what happens when the iron begins to saturate. This problem can be examined with reference to Fig. 4. Flux has been forced out of the iron into the air in the regions of heavy saturation. Whenever flux emerges from iron there must be surface pole strength at that place. Clearly such surface pole strength will be of such a sign as to increase the magnetic force in the part of the tube that is saturated, which is exactly what is needed to keep the flux within the wall of the tube. Since the permeability of iron is so large, a small leakage flux will be sufficient to give a considerable increase in wall flux. We have here a self-relieving mechanism and see why the wall flux depends so little on the permeability. The surface pole strength will have to adjust itself, and there is no longer a pure cosine distribution.

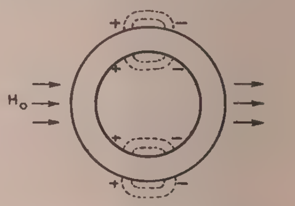


Fig. 4.—Saturation effects.

The reader may suspect that there is a contradiction in the argument. We have previously stated that the surface pole strength depends on the ratio $(\mu - 1)/(\mu + 1)$ and is thus unlikely to vary even with large variations of μ, whilst there must be a variation of surface pole strength to account for the absence of change of the wall flux density.

To resolve this apparent contradiction it must be noted that saturation necessitates an increased magnetic force in the wall of the tube if the flux density is to be maintained constant; but the magnetic force in the wall is the difference between the external magnetic force H_0 and the magnetic force of the induced pole strength $2\pi\mu_0\sigma_a$, neglecting the small induced pole strength $\mu_0\sigma_b$ for the sake of simplicity. Thus the magnetic force in the wall is

$$H_0 - 2\pi\sigma_a = (1 - y)H_0 = \frac{2H_0}{\mu + 1} \quad . \quad . \quad . \quad (22)$$

and this magnetic force is very sensitive to a change in μ. Thus it is seen that only a small change in σ_a is needed to make a very large change in the magnetic force in the wall, and the contradiction is resolved.

Any such change in the surface pole strength will introduce harmonics. In general, the surface pole strength will now be given by

$$\sigma_a = \sigma_{a1}\cos\theta + \sigma_{a3}\cos 3\theta + \sigma_{a5}\cos 5\theta + \ldots \quad . \quad (23)$$

and $$\sigma_b = \sigma_{b1}\cos\theta + \sigma_{b3}\cos 3\theta + \sigma_{b5}\cos 5\theta + \ldots \quad . \quad (24)$$

As a result the field in the space surrounded by the tube will be given by

$$H_r = (H_0 - 2\pi\sigma_{a1} + 2\pi\sigma_{b1})\cos\theta - \left(2\pi\sigma_{a3}\frac{r^2}{a^2} - 2\pi\sigma_{b3}\frac{r^2}{b^2}\right)\cos 3\theta - \ldots \quad . \quad (25)$$

$$-H_\theta = (H_0 - 2\pi\sigma_{a1} + 2\pi\sigma_{b1})\sin\theta - \left(2\pi\sigma_{a3}\frac{r^2}{a^2} - 2\pi\sigma_{b3}\frac{r^2}{b^2}\right)\sin 3\theta - \ldots \quad . \quad (26)$$

Thus the harmonics in the surface polarity distribution will be of the same order as $H_0 - 2\pi\sigma_{a1} + 2\pi\sigma_{b1}$, or $1/\mu \times 2\pi\sigma_{a1}$, and they will be highly dependent on the permeability; it has already been pointed out that the harmonics only arise as a result of saturation. The conclusion is that the induced surface pole strength on the iron tube will consist of a fundamental cosine term which will be largely independent of μ and of the order of the applied field, and of a series of odd harmonics which will depend directly on μ and will be of the order of $1/\mu$ times the applied field. We have now to consider the effect of this distribution of surface pole strength on the tube and on the space outside or inside the tube.

Consider first the region surrounding the tube. The radial field will be given by

$$H_r = \left[H_0 + 2\pi\sigma_{a1}\left(\frac{a}{r}\right)^2\right]\cos\theta + 2\pi\sigma_{a3}\left(\frac{a}{r}\right)^4\cos 3\theta + \ldots \quad . \quad . \quad . \quad . \quad (27)$$

if the effect of the pole strength on the inner wall is neglected for simplicity. Now we know that σ_{a1} is of the order of H_0 and σ_{a3} is of the order of $\frac{1}{\mu}\sigma_{a1}$. Hence the harmonics can have no possible effect so long as $\mu > 100$, say. If we explore the field surrounding the tube at various states of saturation, we shall find a flux distribution varying as $\cos\theta$ and independent of μ.

Consider next the region occupied by the iron of the tube where

$$B_\theta \simeq \frac{2a\mu_0 H_0}{t}\sin\theta$$

as already shown, and thus the effect of saturation is likely to be very small until the iron is heavily saturated, so long as $\mu \gg \frac{a+b}{t}$.

Lastly let us consider the field inside the space surrounded by the tube, which depends [eqns. (10) to (14)] dominantly on the permeability of the iron and thus on the harmonics in the surface pole strength. Thus the screening effect will depend on these harmonics.

It has already been shown that, as long as μ is constant, the field in the screened space is given by

$$H_i = H_0\left(\frac{1 - y^2}{1 - y^2x^2}\right) \simeq H_0\frac{2a}{\mu t} \quad . \quad . \quad . \quad (28)$$

as long as $a \simeq b$ and $\mu \gg \frac{a+b}{t}$.

If we consider the case of a 3 in-diameter tube of thickness 0·036 in, $\mu \gg 83$. While this condition is not difficult to obtain with Mumetal at low flux densities, it is essential to see the magnitude of the limit imposed. Since μ can vary considerably around the wall of the tube (from say 60 000 to 5 000) and still keep within the limit of $\mu \gg 83$, it is natural to ask which value of μ should be taken to estimate the screening effect for a particular tube.

It is necessary to consider the magnetic field due to the harmonics in the pole-strength distribution. For the third harmonic ($2\pi\mu_0\sigma_3\cos 3\theta$), it is clear that the field will resemble that of a 6-pole machine. At the centre of the tube such a distribution of pole strength can never produce any field, because there will be equal numbers of north and south poles at equal distances and their effect will cancel out. But there will be a field at every other point inside the tube, and this will become stronger nearer to the iron surface. The general expression for the magnetic field due to the nth harmonic is given by

$$H_r = -\,2\pi\sigma_n\left(\frac{r}{R}\right)^{n-1}\cos n\theta \quad . \quad . \quad (29)$$

and $$H_\theta = 2\pi\sigma_n\left(\frac{r}{R}\right)^{n-1}\sin n\theta \quad . \quad . \quad . \quad . \quad (30)$$

where (r, θ) are the co-ordinates of a point at which the field is measured and R is the radius of the shell of pole strength $\mu_0\sigma_n$. At the centre of the tube the field must therefore depend entirely on the fundamental component of polarity, since the higher harmonics give zero field at this point. If harmonics are present in the surface-polarity distribution, it is no longer possible to refer to the screening ratio of the tube without reference to a particular point therein, since the screening ratio will vary from point to point. The effect of the harmonics will be a maximum when r is maximum, i.e. at the inner surface of the tube. For the centre of the tube, we can now decide on the value of μ to be used in the expression for $\mu t/2a$ the screening ratio $H_0/H_i = \mu t/2a$. The correct value is that corresponding to the fundamental component of surface pole strength.

Consider this with reference to Fig. 5, in which harmonics higher than the third have been neglected. Clearly the magnitude of the fundamental component of H_θ is governed largely by the magnitude of H_θ required at angles near 90°, i.e. near the place at which μ is a minimum. But it would be unduly pessimistic to choose this minimum value of μ as being correct for the screening ratio. In the example illustrated in Fig. 5 the magnitude of H_θ should be found at $\theta = 60°$ since the third harmonic will be zero at this angle. The magnitude of the fundamental can then be derived, and will give the correct permeability. If other harmonics are also prominent, it may be necessary to carry out a harmonic analysis of the H_θ curve and thus to obtain its fundamental component.

If it is desired to find the screening ratio for points other than the centre of the tube, the effect of the harmonics will have to

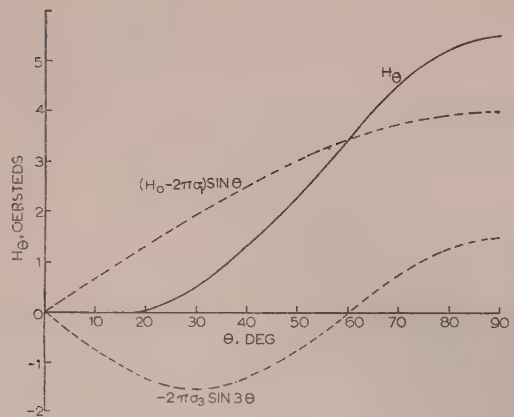


Fig. 5.—Variation of magnetic force in tube wall.

be considered. If the discussion is again confined to the third harmonic, it is clear that it must reinforce the field at $\theta = 90°$ (see Fig. 5) and must be of the same sign as $H_0 - 2\pi\sigma_1$ at $\theta = 90°$. Hence σ_1 will be of the same sign as σ_3, and the field strength will be

$$(H_0 - 2\pi\sigma_1) \sin \theta - 2\pi\sigma_3 \sin 3\theta$$

at points close to the inner wall of the tube.

(3) EXPERIMENTAL INVESTIGATION

A substantially uniform magnetic field was produced by passing a current through two coaxial coils each having 2250 turns. The mean diameter of the coils was 21 cm, and their mid-planes were 38 cm apart. The field was explored by means of a search coil of 600 turns of No. 44 s.w.g. enamelled wire, the cross-section of the coil being approximately 5 cm × 1 cm. It was found that the magnetic flux density in a region half-way between the magnetizing coils and on their axis was $B = 30$ lines/cm^2/amp. The heat dissipation of the magnetizing coils and the rupturing capacity of the reversing switch connected in their circuit limited the current to 4 amp. Hence the maximum flux density was 120 lines/cm^2.

To examine the screening effect of iron tubes with widely different amounts of saturation, it was decided to test a mild-steel tube which was unsaturated and a nickel iron tube which could be very heavily saturated. The screening ratio was obtained by means of a ballistic galvanometer connected to the search coil described above, and also by means of a.c. tests when the search coil was connected to a cathode-ray oscillograph through an integrating circuit using a pentode valve.

(3.1) Screening Effect of Mild Steel Tube

The length of the tube was 15·3 cm, the mean diameter 7·15 cm, and the thickness 0·32 cm. The tube was placed transversely to the magnetic field at the centre of the magnetizing system. A search coil of 36 turns was wound on one portion of the wall of the tube, and the tube could be rotated on a graduated scale so as to place the search coil at various angles to the applied magnetic field. By this means, curves of wall flux against angle could be obtained, when different magnetizing currents were reversed by means of the reversing switch. Fig. 6(*a*) shows such a family of curves. It will be seen that the curves are substantially sinusoidal as predicted in Section 2.2, and the wall flux is accordingly largely independent of the variation of permeability around the tube. The theoretical values for B at $\theta = 90°$, as calculated for a tube of infinite length, are compared with the actual values of B in Table 1. The difference between the measured and calculated values of wall flux density is probably due to the fact that the field was not strictly uniform, and more important still, the tube was relatively short, its length being only about twice the diameter. The constancy of the ratio shows the soundness of the analysis, when applied to an iron tube of varying permeability.

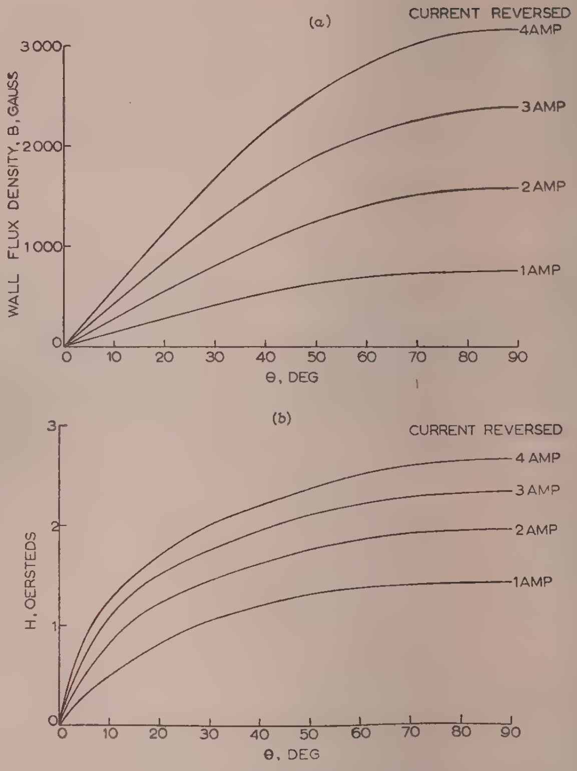


Fig. 6.—Wall of mild-steel tube.

(*a*) Flux density.
(*b*) Magnetic force.

Table 1

COMPARISON OF THEORETICAL AND ACTUAL VALUES

Current	B_0	Wallflux density B		Ratio of experimental to theoretical
		Theoretical	Experimental	
amp	gauss	gauss	gauss	
1	30	670	750	1·12
2	60	1 340	1 575	1·175
3	90	2 010	2 380	1·18
4	120	2 680	3 150	1·175

To examine the effects of surface polarity it was necessary to obtain curves of magnetizing force against angle by obtaining a B/H reversal curve for the tube, which was done in the usual manner. A magnetizing winding was wound on the tube, and the flux density was measured by means of a fluxmeter connected to the search coil on the wall of the tube. Fig. 7 shows the B/H and μ/H curves obtained. It is seen that the tube is very

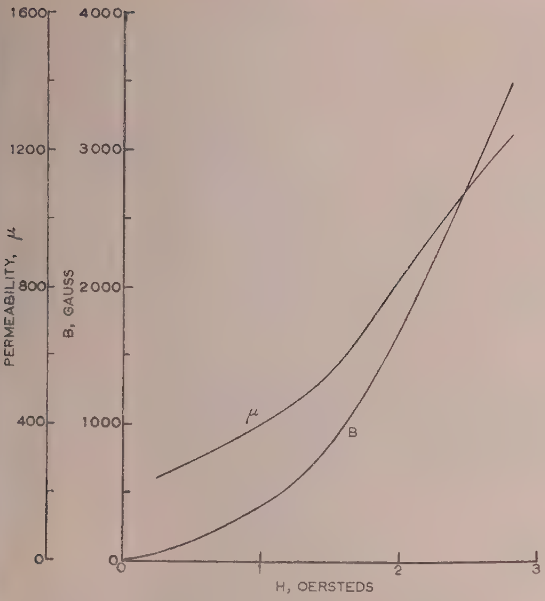


Fig. 7.—Magnetic reversal curve for mild-steel tube.

unsaturated in the range of flux densities covered by the test, but that there is a variation of permeability of about 6 : 1 because of the low permeabilities at very weak field strengths. Fig. 6(*b*) shows curves of H/θ, where H is the magnetizing force in the tube wall and θ is the angle around the circumference of the tube. Comparison of Figs. 6(*a*) and 6(*b*) shows that, while the flux density varies sinusoidally around the tube, the magnetizing force does not, but consists, as predicted, of a fundamental distribution with various harmonics superimposed on it. This clearly shows that the surface polarity has harmonics in its distribution. The fundamentals of the H/θ curves in Fig. 10 are given in Table 2.

Table 2

FUNDAMENTALS OF H/θ CURVES

Current	Fundamental H	Maximum H
amp	oersteds	oersteds
1	1·6	1·42
2	2·22	1·95
3	2·65	2·32
4	3·02	2·65

It was pointed out in Section 2.2 that harmonics in the distribution of surface polarity around the tube would result in a magnetic field that varied from place to place in the space surrounded by the tube. Thus the screening ratio will also vary. The internal field was measured in three positions (see Fig. 8) by the search coil connected to a ballistic galvanometer. Fig. 9(*a*) shows the screening ratios obtained for different magnetizing currents. For comparison the curves of theoretical screening ratio for positions A and B have been plotted in Fig. 9(*a*). These theoretical curves are obtained from eqn. (14) $H_0/H_i = \mu t/2a$. The permeability chosen for the screening ratio at position B is that corresponding to maximum H, while the permeability for position A is that corresponding to the fundamental component of H, i.e. to that value of μ which occurs at the point where the H/θ curve crosses the curve of fundamental H against θ. Fig. 9(*a*) shows clearly that the highest screening ratio for an unsaturated iron tube is to be expected at the position B. This is a result of considerable interest, since static screening is often required from weak magnetic fields. Table 3 shows the ratio of the theoretical to experimental screening ratio. Comparison with Table 1 shows that the difference must again be due largely to the end effect. It is clear that the theoretical treatment can be used to give a very close approximation to the experimental screening ratio.

Fig. 9(*a*) shows the screening ratios obtained on an a.c. test

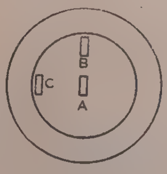


Fig. 8.—Positions of search coil in tube.

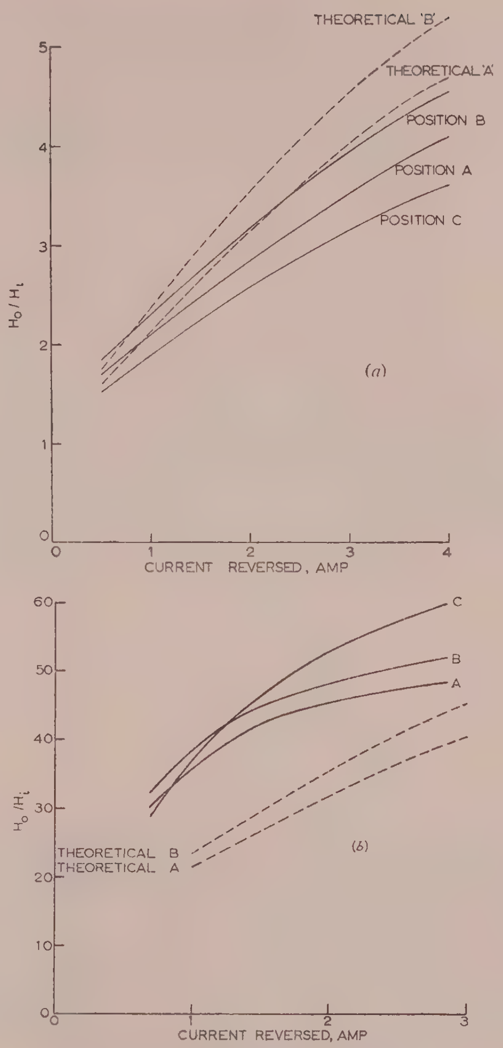


Fig. 9.—Screening ratio of mild-steel tube.
(*a*) D.C. test.
(*b*) A.C. test.

Table 3

Comparison of Theoretical and Experimental Screening Ratios

Current	Experimental screening ratio		Theoretical screening ratio		Ratio of theoretical to experimental screening ratio	
	A	B	A	B	A	B
amp						
1	21	23	21·3	23·3	1·01	1·01
2	28·5	32	31·5	35·5	1·1	1·11
3	35·5	39·5	40·5	45·5	1·14	1·15
4	41	45·5	47	53	1·15	1·16

carried out at a frequency of 50 c/s. The screening effect is considerably improved. The screening due to current flowing up over half the section of the tube and down over the other half was negligible owing to the high resistance of the tube. The improvement of the screening ratio must therefore be due to eddy currents flowing in one direction on the inside of the tube and returning on the outside of the tube (see Fig. 10). These eddy currents will crowd the flux into the surface layers of tube wall, and this will result in an increase in permeability because the tube is so unsaturated. It is of interest to note that the full benefit of this increased permeability is felt only at C, while at A and B there is a smaller increase in the screening ratio because of the field of the eddy currents.[4] The criterion of eddy-current effect, mt, given in the Reference is approximately 4, so that the eddy-current effect is likely to be very pronounced, as is the case. The theoretical d.c. curves are plotted in Fig. 9(*b*) for comparison.

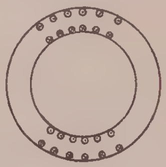


Fig. 10.—Eddy currents in tube.

(3.2) Screening Effect of Mumetal Tube

The length of the tube was 15·3 cm, the mean diameter 7·7 cm and the thickness 0·09 cm. A search coil of 36 turns was wound on one portion of the wall of the tube. Fig. 11 shows

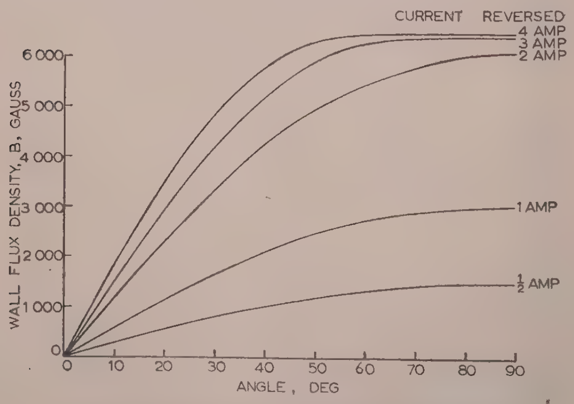


Fig. 11.—Flux density in wall of Mumetal tube.

Table 4

Comparison of *B* Values

Current	B_0	Wall flux density *B*		Ratio
		Theoretical	Experimental	
amp	gauss	gauss	gauss	
½	15	1 280	1 500	1·17
1	30	2 570	3 020	1·17
2	60	5 140	6 100	1·19
3	90	7 700	6 410	0·835
4	120	10 300	6 500	0·63

a family of curves of wall flux density *B* plotted against θ, the angle measured around the circumference of the tube. The theoretical values for *B* at $\theta = 90°$ are compared with the actual values in Table 4.

The Table shows that the ratio of experimental to theoretical

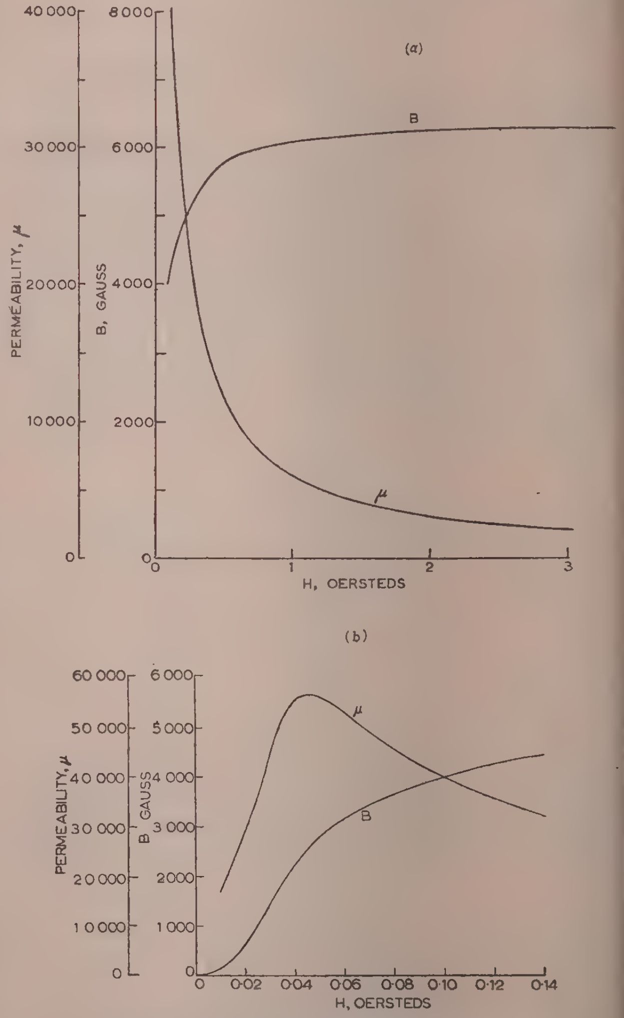


Fig. 12.—Magnetic reversal curve for Mumetal tube

(*a*) High flux density.
(*b*) Low flux density.

flux density is the same for the Mumetal tube, when unsaturated, as for the mild-steel tube, in spite of the fact that the thickness and permeability of the tubes are widely different. This underlines the usefulness of the theory in predicting the behaviour of any iron tube.

Figs. 12(*a*) and 12(*b*) give the magnetic reversal curve of the Mumetal tube, Fig. 12(*b*) showing the part of this curve at low flux densities. It will be seen that there is a slight inconsistency between Figs. 11 and 12(*a*), since Fig. 11 shows the saturated flux density of 6 500 gauss and Fig. 12(*a*) 6 300 gauss. This inconsistency is due to lack of homogeneity of the tube. The examination of H/θ curves had therefore to be confined to those obtained with magnetizing currents smaller than 3 amp. Figs. 13(*a*) and 13(*b*) show curves of H/θ for the Mumetal tube.

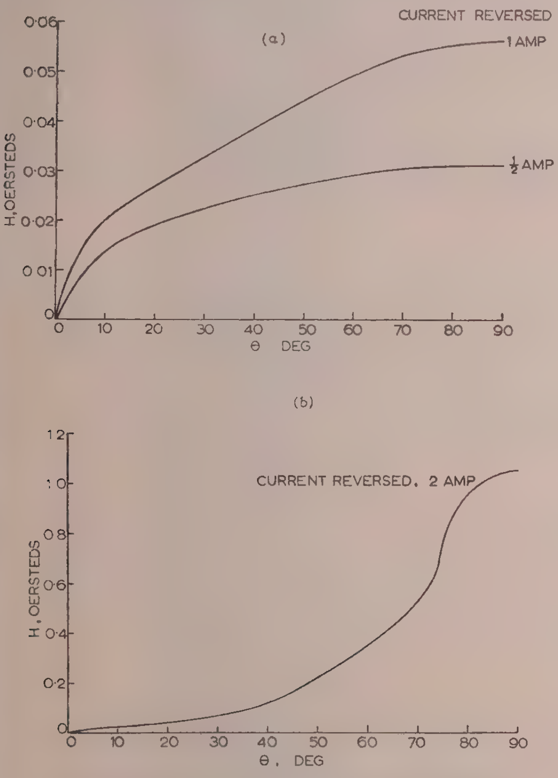


Fig. 13.—Magnetic force in wall of Mumetal tube.

(*a*) Low flux density.
(*b*) High flux density.

The curve of H/θ taken with a magnetizing current of ½ amp is of the same shape as the curves obtained from the mild-steel tube, while the curve obtained at a magnetizing current of 2 amp is completely different owing to the onset of saturation in the Mumetal. The first three terms of the Fourier analysis for these two curves are as follows:

Current	*Magnetizing force*
½ amp	$H = 0{\cdot}0348 \sin\theta + 0{\cdot}0055 \sin 3\theta + 0{\cdot}0026 \sin 5\theta$
2 amp	$H = 0{\cdot}6055 \sin\theta - 0{\cdot}3061 \sin 3\theta + 0{\cdot}1173 \sin 5\theta$

It can be seen that, with increasing saturation, the harmonics have become more pronounced and that the third harmonic has changed its sign. When the third harmonic has the negative sign it will add to the internal field at $\theta = 90°$, and the position B will now give the lowest screening ratio and position C the highest. This is seen from Fig. 14(*a*), which shows the screening ratio obtained with the Mumetal tube with heavy saturation [Fig. 14(*a*) should be compared with Fig. 9(*a*)]. It was not possible to obtain measurements on the ballistic galvanometer when the current reversed was less than 2½ amp. Because of this lack of sensitivity and the inhomogeneity of the Mumetal tube, no theoretical screening-ratio curve has been plotted on Fig. 14(*a*). The theoretical values of the screening ratio at 2 amp obtained from the formula $H_0/H_i = \mu t/2a$ are as follows:

Position	A	B
Screening ratio	120	68

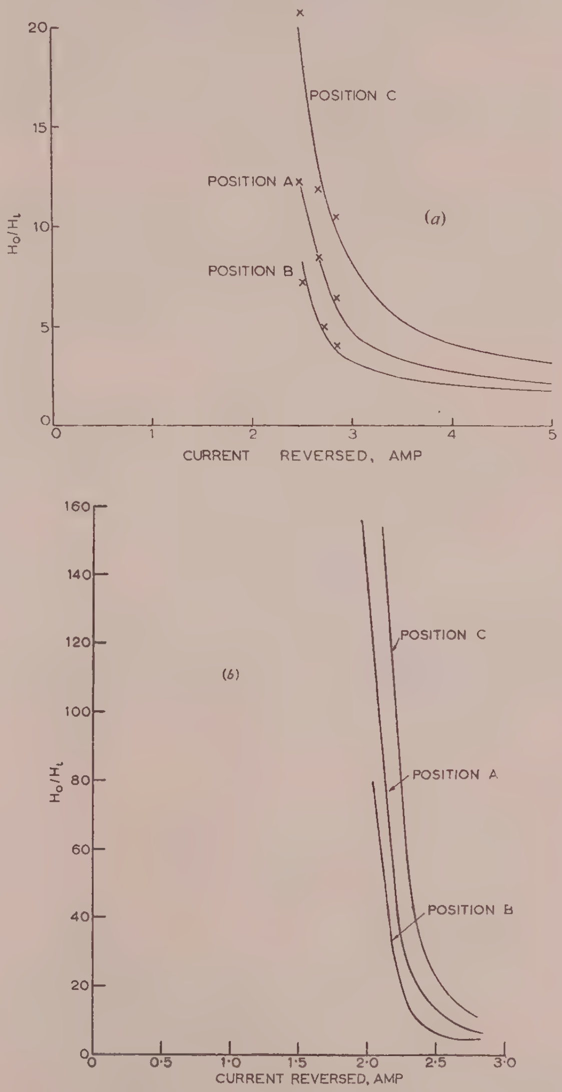


Fig. 14.—Screening ratio of Mumetal tube.

(*a*) D.C. test.
A.C. readings . . . ×
(*b*) A.C. test.

Fig. 14(*b*) shows the screening ratio of the Mumetal tube obtained from a.c. measurements. Some of these a.c. readings have been inserted in Fig. 14(*a*) for comparison, and show that the a.c. and d.c. curves practically coincide. This is to be expected from a consideration of the eddy-current criterion, which in the case of the Mumetal tube will be only about $mt = 1 \cdot 5$. The eddy-current effect will therefore be very small. Examination of Figs. 14(*a*) and 14(*b*) shows that the magnetic field in a saturated iron tube will vary considerably from place to place. Clearly it is impossible to speak of the screening effect of such a tube without reference to the point concerned. Figs. 15(*a*), 15(*b*) and 15(*c*) show oscillograph traces of the flux density outside the tube, in the tube wall and in the space surrounded by the tube. These were obtained with the Mumetal tube in an alternating magnetic field, the alternating exciting current being approximately 3 amp (r.m.s.). These traces show very clearly the type of variation predicted in Section 2.2, namely that the external field and wall flux are independent of μ, but that the field in the space surrounded by the tube depends dominantly on μ.

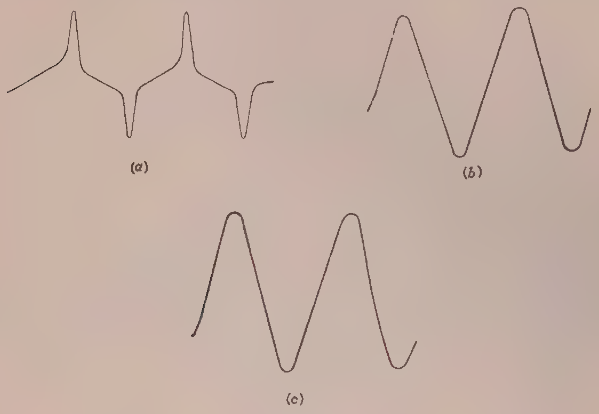


Fig. 15.—Flux density with Mumetal tube.

Exciting current, 3⅛ amp (r.m.s.).
(*a*) Inside.
(*b*) Outside.
(*c*) In tube wall.

(4) CONCLUSION

It has been shown that the magnetic behaviour of iron of widely varying permeability can be accurately predicted from a theory based on the assumption of constant permeability, provided that a close study is made of the distribution of surface polarity. The theory has been applied to the screening action of iron tubes, and it has been shown that the screening effect will vary from place to place in the screened region. This variation has been measured and has been found to agree closely with the theoretical predictions. Thus accurate estimation is possible of the screening effect of such tubes.

(5) ACKNOWLEDGMENT

The author gratefully acknowledges the guidance and encouragement given by Prof. E. B. Moullin.

(6) REFERENCES

(1) HAMMOND, P.: "Leakage Flux and Surface Polarity in Iron Ring Stampings," *Proceedings I.E.E.*, Monograph No. 116, January, 1955 (**102** C, p. 138).

(2) MOULLIN, E. B.: "Principles of Electromagnetism" (Oxford University Press, 1955), p. 356.

(3) JEANS, SIR JAMES: "Electricity and Magnetism" (Cambridge University Press, 1927), para. 416.

(4) Reference 2, Fig. 154.

621.372.8:662.3.012.8

Monograph No. 145 R
Aug. 1955

THE CALCULATION OF THE EQUIVALENT CIRCUIT OF AN AXIALLY UNSYMMETRICAL WAVEGUIDE JUNCTION

By R. E. COLLIN, B.Sc., Ph.D., and JOHN BROWN, M.A., Ph.D., Associate Member.

(*The paper was first received 24th February, and in revised form 4th May,* 1955. *It was published as an* INSTITUTION MONOGRAPH *in August,* 1955.)

SUMMARY

The parameters of the equivalent circuit of an axially unsymmetrical waveguide junction may be calculated by a direct application of the variational method developed by Schwinger. Certain useful properties of this method exist only when the junction has a certain degree of symmetry. A modification to the method by which these properties are retained even when the junction is asymmetrical is developed and shown to be closely related to the Weissfloch method for the experimental determination of the equivalent-circuit parameters. The method is illustrated by numerical calculations for the junction between an empty waveguide and one partially filled with dielectric: good agreement is found with experimental results.

LIST OF PRINCIPAL SYMBOLS

x, y, z = Cartesian co-ordinates.
E_x = x-component of electric field.
H_y, H_z = y- and z-components of magnetic field.
a, b = Dimensions of waveguide cross-section.
t = Thickness of dielectric slab.
ϵ_r = Relative permittivity of dielectric slab.
Z_1, Z_2 = Characteristic impedances of transmission lines in equivalent circuit.
n = Turns ratio of ideal transformer.
u, v = Additional lengths of transmission lines in equivalent circuit.
β_0 = Free-space phase-change coefficient.
Y_0 = Free-space wave admittance.
β_{01} = Phase-change coefficient for H_{01} mode in empty waveguide.
α_{0n} = Attenuation coefficient for H_{0n} mode in empty waveguide ($n \geqslant 2$).
Y_{0n} = Wave admittance for H_{0n} mode.
β_1 = Phase-change coefficient for dominant mode in partially filled guide.
α_n = Attenuation coefficients for evanescent modes in partially filled waveguide ($n \geqslant 2$).
Y_n = Wave admittances for modes in partially filled waveguide function defining field variation of nth mode ($n \geqslant 1$).
A = Amplitude of incident H_{01} mode.
A_n = Amplitude of H_{0n} mode reflected by junction ($n \geqslant 1$).
B_n = Amplitude of nth mode in partially filled guide ($n \geqslant 1$).
$F(y), G(y)$ = Functions defining electric and magnetic fields respectively in the plane $z = 0$.
Y_A = Input admittance of junction.
μ = Absolute permittivity of free-space.
γ = Propagation coefficient.

(1) INTRODUCTION

Any loss-free junction between two waveguides may be represented by an equivalent circuit involving at most three unknown parameters, provided that each waveguide can support only one mode of propagation.[1] Each waveguide is represented by a transmission line in the sense that the voltage and current at any point P_1 on the transmission line (Fig. 1) are proportional to the electric and magnetic field strengths of the propagated mode at a corresponding point P on the axis of the waveguide. The distances of P and P_1 from their respective origins O and O_1 are equal, and this requires that the phase-change coefficient of the transmission line and the propagated mode in the waveguide should be equal. The characteristic impedance of the equivalent transmission line is not uniquely defined and may, subject to the possible restriction discussed below, be made to have any arbitrarily selected value.

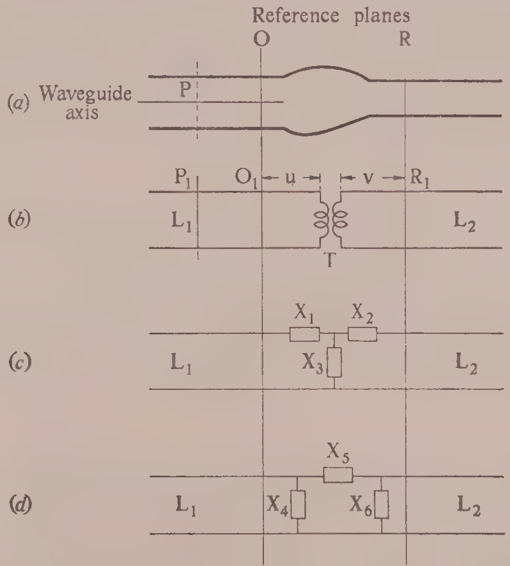


Fig. 1.—Equivalent circuit of a waveguide junction.

(*a*) Cross-section of the junction.
(*b*), (*c*) and (*d*) Possible equivalent circuits.
L_1, L_2 are transmission lines with phase coefficients equal to those of the propagated modes in the corresponding waveguides: their characteristic impedances are Z_1 and Z_2 respectively.
T is an ideal transformer of turns ratio $n : 1$.
X_1, X_2, etc., are reactances.

The three parameters specifying the junction may be chosen in a variety of ways, several being illustrated by Fig. 1. The methods by which the various forms of equivalent circuit may be transformed into each other have been fully discussed by Marcuvitz.[2] The choice of equivalent circuit depends on such factors as the complexity of the frequency dependence of its parameters, and must be based on the properties of the particular junction which is being considered. For general studies, how-

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Collin, who was formerly in the Department of Electrical Engineering at the Imperial College of Science and Technology, is now at C.A.R.D.E., Quebec.
Dr. Brown is in the Department of Electrical Engineering, Imperial College of Science and Technology.

ever, the circuit shown in Fig. 1(*b*) is particularly convenient since it forms the basis of a method by which the parameters may be determined experimentally.[3] The three parameters for this circuit are *u* and *v*—the lengths of the transmission lines inserted between the reference planes O_1 and R_1—and *n*, the turns ratio of the ideal transformer, provided that the characteristic impedances of the transmission lines, Z_1 and Z_2 are specified. An alternative is to specify only one characteristic impedance, say Z_1, to make the ideal transformer of unity turns ratio and then to use the ratio Z_2/Z_1 as the third parameter. This forms the restriction on the choice of characteristic impedance referred to in the previous paragraph.

Exact values for the parameters can be obtained only in a few simple cases and a number of methods by which approximate values may be obtained have been developed. One of the most powerful is a variational technique which has been developed by Schwinger.[4] In Schwinger's own applications of this technique the junctions considered have sufficient symmetry to reduce the required number of parameters to one or two. This leads to certain desirable features, one of which is that the method can be used to indicate the maximum error which can result from the approximations made. The variational technique can be used for axially unsymmetrical junctions requiring three parameters as has been demonstrated by Miles[5] and Lewin,[6] but no indication of the error is forthcoming from their results. Furthermore, complex functions have to be used, whereas Schwinger's original work involved only real functions.

A modification to the variational method has been tried and enables axially unsymmetrical junctions to be examined, while preserving the advantages of indicating the maximum error in the results and of requiring only real functions. The modified approach is closely related to the method by which the junction parameters can be obtained experimentally, and is best described by reference to an actual example, that chosen being the junction between an empty waveguide and one partially filled with dielectric.

(2) NATURE OF THE FIELD NEAR THE INTERFACE

The waveguide junction being considered is shown in Fig. 2. A rectangular waveguide, having cross-sectional dimensions *a* and *b*, is empty for negative values of *z* and partially filled by a slab of dielectric of thickness, *t*, for positive values of *z*. The dimensions, *a* and *b*, are such that only the H_{01} mode can propagate in the empty guide: the principal part of the field for negative values of *z* is therefore given by the equations

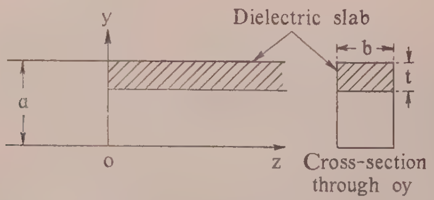


Fig. 2.—Junction between an empty waveguide and one partially filled with dielectric.

$$E_x = (A\varepsilon^{-j\beta_{01}z} + A_1\varepsilon^{j\beta_{01}z})\sin(\pi y/a) \quad . \quad . \quad . \quad (1)$$

$$H_y = Y_{01}(A\varepsilon^{-j\beta_{01}z} - A_1\varepsilon^{j\beta_{01}z})\sin(\pi y/a) \quad . \quad . \quad (2)$$

where A = Amplitude of the wave incident on the junction.
A_1 = Amplitude of the reflected wave.
β_{01} = Phase-change coefficient of the H_{01} mode.
and Y_{01} = Wave admittance of the H_{01} mode.

The last two quantities are given by the equations

$$\beta_{01} = [\beta_0^2 - (\pi/a)^2]^{1/2} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$Y_{01} = \beta_{01}Y_0/\beta_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

There is also an H_z field component associated with the H_{01} mode, but its value is not required.

The nature of the discontinuity at the plane $z = 0$, is such that only the field components E_x, H_y and H_z are excited anywhere in the structure and that these components are all independent of the co-ordinate *x*. Subject to certain restrictions on the relative permittivity and the thickness of the dielectric slab, only one mode propagates within the partially filled waveguide. The phase-change coefficient, β_1, the wave admittance, Y_1, and the field distribution for this propagated mode are derived in Section 10.1. The principal part of the field for positive values of *z* is therefore

$$E_x = B_1\phi_1(y)\varepsilon^{-j\beta_1 z} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$H_y = B_1Y_1\phi_1(y)\varepsilon^{-j\beta_1 z} \quad . \quad . \quad . \quad . \quad (6)$$

where B_1 is the amplitude coefficient and $\phi_1(y)$ is a function which specifies the field variation in a plane at right angles to the directon of propagation.

The fields existing with the complete waveguide structure must be such that the tangential components of the electric and magnetic field strengths, E_x and H_y respectively, must be continuous over the interface plane, $z = 0$. Since the function $\phi_1(y)$ does not equal $\sin(\pi y/a)$, the fields described by eqns. (1), (2), (5) and (6) cannot form a complete solution by themselves. An additional field must exist and takes the form of a sum of evanescent modes in each of the regions $z \leqslant 0$ and $z \geqslant 0$. In the former these evanescent modes are of H_{0n} type, for which the fields are

$$E_x = A_n\sin(n\pi y/a)\varepsilon^{\alpha_{0n}z} \quad . \quad . \quad . \quad . \quad . \quad (7)$$

$$H_y = -A_nY_{0n}\sin(n\pi y/a)\varepsilon^{\alpha_{0n}z} \quad . \quad . \quad . \quad (8)$$

n taking all values from 2 to infinity. Only terms which decay exponentially in the direction away from the junction can be excited, so that terms involving $\varepsilon^{-\alpha_{0n}z}$ are excluded. The attenuation coefficient, α_{0n}, and wave admittance Y_{0n} are given by

$$\alpha_{0n} = [(n\pi/a)^2 - \beta_0^2]^{1/2} \quad . \quad . \quad . \quad . \quad (9)$$

$$Y_{0n} = -j\alpha_{0n}Y_0/\beta_0 \quad . \quad . \quad . \quad . \quad . \quad (10)$$

all the α_{0n} being real if $2\pi/a$ exceeds β_0.

A similar set of evanescent modes is excited for positive values of *z*, and the fields for these are

$$E_x = B_n\phi_n(y)\varepsilon^{-\alpha_n z} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

$$H_y = B_nY_n\phi_n(y)\varepsilon^{-\alpha_n z} \quad . \quad . \quad . \quad . \quad (12)$$

only fields decaying in the positive *z* direction being excited. B_n is the amplitude constant for the *n*th mode, *n* taking all values from 2 to infinity. The attenuation coefficients, α_n, the wave admittances Y_n and the functions $\phi_n(y)$ are derived in Section 10.1. The functions $\phi_n(y)$ are normalized to satisfy the equation

$$\left.\begin{aligned}\int_0^a \phi_n(y)\phi_m(y)dy &= 0 \text{ if } m \neq n\\ &= 1 \text{ if } m = n\end{aligned}\right\} \quad . \quad . \quad . \quad (13)$$

for all values of *m* and *n* from 1 to infinity.

General expressions for the fields throughout the waveguide structure may now be written down.

'hen $z \leqslant 0$

$${}_x = (A\varepsilon^{-j\beta_{01}z} + A_1\varepsilon^{j\beta_{01}z}) \sin(\pi y/a) + \sum_{n=2}^{\infty} A_n \sin(n\pi y/a)\varepsilon^{+\alpha_{0n}z} \quad . \quad . \quad . \quad . \quad . \quad (14)$$

$${}_y = Y_{01}(A\varepsilon^{-j\beta_{01}z} - A_1\varepsilon^{j\beta_{01}z}) \sin(\pi y/a) - \sum_{n=2}^{\infty} Y_{0n}A_n \sin(n\pi y/a)\varepsilon^{\alpha_{0n}z} \quad . \quad (15)$$

ıd when $z \geqslant 0$

$$E_x = B_1\phi_1(y)\varepsilon^{-j\beta_1 z} + \sum_{n=2}^{\infty} B_n\phi_n(y)\varepsilon^{-\alpha_n z} \quad . \quad . \quad (16)$$

$$H_y = Y_1B_1\phi_1(y)\varepsilon^{-j\beta_1 z} + \sum_{n=2}^{\infty} Y_nB_n\phi_n(y)\varepsilon^{-\alpha_n z} \quad . \quad (17)$$

ı the above equations, the constants A_n and B_n are unknown ıd must be calculated to satisfy the continuity of E_x and H_y at $= 0$. As in most waveguide problems of this type, there is :tle prospect of obtaining an exact solution because of the omplexity of the equations. It may be noted at this point that nly the values of A_1 and B_1 are needed to calculate the equivalent-rcuit parameters.

(3) THE VARIATIONAL METHOD OF SOLUTION

In eqns. (14)–(17) the waveguide fields are expressed in rms of the unknown amplitude coefficient. The first step in oplying the variational method is to express the fields in terms f the tangential electric field in the junction plane, $z = 0$: ıppose this field is given by the function F(y), which is, of course, ot known until the complete solution has been obtained. From ıns. (14) and (16),

$$(A + A_1) \sin(\pi y/a) + \sum_{n=2}^{\infty} A_n \sin(n\pi y/a) = \mathrm{F}(y) \quad . \quad (18)$$

$$\sum_{n=1}^{\infty} B_n\phi_n(y) = \mathrm{F}(y) \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

ach of these equations is valid when y lies between 0 and a, ıd together they imply the continuity of E_x at the plane $z = 0$. Eqn. (18) is an ordinary Fourier series, so that

$$A + A_1 = \frac{2}{a}\int_0^a \mathrm{F}(y') \sin(\pi y'/a)dy' \quad . \quad . \quad . \quad (20)$$

$$A_n = \frac{2}{a}\int_0^a \mathrm{F}(y') \sin(n\pi y'/a)dy' \quad (n = 2, 3, \ldots) \quad . \quad (21)$$

qn. (19) has a form similar to a Fourier series and the orthoonal properties of the functions $\phi_n(y)$ [eqn. (13)] lead to the sult

$$B_n = \int_0^a \mathrm{F}(y')\phi_n(y')dy' \quad (n = 1, 2, \ldots) \quad . \quad . \quad (22)$$

Substitution for A_n and B_n in eqns. (15) and (17) and the condion that H_y is continuous for $z = 0$ gives

$$Y_{01}(A - A_1) \sin(\pi y/a) - \frac{2}{a}\sum_{n=2}^{\infty} Y_{0n}\int_0^a \mathrm{F}(y') \sin(n\pi y'/a) \sin(n\pi y/a)dy'$$
$$= \sum_{s=1}^{\infty} Y_s\int_0^a \mathrm{F}(y')\phi_s(y')\phi_s(y)dy' \quad (0 \leqslant y \leqslant a) \quad . \quad (23)$$

This is an integral equation for the unknown function F(y) [the constant A_1 on the left may be expressed in terms of F(y) by eqn. (20), but the above form is more convenient for the later analysis]. A solution of the integral equation is just as difficult to obtain as a solution for the constants A_n, B_n, but the equation can be used to give an expression for A_1 giving quite accurate results if a relatively crude approximation for F(y) is known. To obtain this expression multiply both sides of eqn. (23) by F(y) and integrate with respect to y from 0 to a.

Thus:

$$Y_{01}(A - A_1)\int_0^a \mathrm{F}(y)\sin(\pi y/a)dy = \sum_{s=1}^{\infty} Y_s\left[\int_0^a \mathrm{F}(y')\phi_s(y')dy'\right]^2 + \frac{2}{a}\sum_{n=2}^{\infty} Y_{0n}\left[\int_0^a \mathrm{F}(y') \sin(n\pi y'/a)dy'\right]^2 \quad . \quad (24)$$

Now divide both sides by $\left[\int_0^a \mathrm{F}(y) \sin(\pi y/a)dy\right]^2$ and use eqn. (20).

$$\frac{Y_{01}(A - A_1)}{(A + A_1)} = \frac{\frac{a}{2}\sum_{s=1}^{\infty} Y_s\left[\int_0^a \mathrm{F}(y')\phi_s(y')dy'\right]^2 + \sum_{n=2}^{\infty} Y_{0n}\left[\int_0^a \mathrm{F}(y') \sin(n\pi y'/a)dy'\right]^2}{\left[\int_0^a \mathrm{F}(y') \sin(\pi y'/a)dy'\right]^2} \quad . \quad (25)$$

This equation is the basic result from which the Schwinger variation technique proceeds.[4] The left-hand side may be rewritten $Y_{01}(1 - R)/(1 + R)$, where R is the reflection coefficient for the H_{01} mode defined at the plane $z = 0$, and is therefore the input admittance in the equivalent circuit at the point corresponding to $z = 0$, i.e. A in Fig. 3, provided that the characteristic admittance of the line corresponding to the H_{01} mode is selected as Y_{01}. In the equivalent circuit, this input admittance is

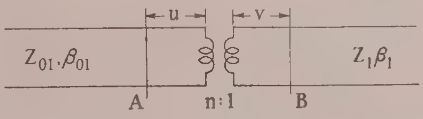


Fig. 3.—Equivalent circuit of the junction shown in Fig. 2.

$$Y_A = Y_{01}\frac{(Y_1/n^2) + jY_{01}\tan(\beta_{01}u)}{Y_{01} + j(Y_1/n^2)\tan(\beta_{01}u)} \quad . \quad . \quad . \quad (26)$$

if the line corresponding to the propagated mode in the partially filled waveguide has the characteristic admittance Y_1. Eqn. (25) therefore enables the input admittance Y_A to be calculated from the electric-field distribution F(y) and then two of the parameters of the equivalent circuit, the turns ratio, n, and the length, u, of the first transmission line may be found from eqn. (26).

The most important property of eqn. (25) is that the expression on the right is stationary with respect to variations of F(y) about the correct function, i.e. the one which satisfies eqn. (23). This means that if an approximation to F(y), differing from the true function by the first order of small quantities, is available, then the value of the input admittance calculated from eqn. (25) will be in error only by a second-order quantity. This stationary property is proved in Section 10.2.

The function F(y) which satisfies eqn. (23) and hence makes Y_A stationary is a complex one, and this considerably complicates the computation of Y_A for approximations to the true function. Since Y_A itself is in general a complex number, it is neither a maximum nor a minimum when the correct function F(y) is used on the right-hand side of eqn. (25). In Schwinger's formulation of the variational method for junctions having a certain amount of symmetry, he contrived to obtain an expression similar in form to eqn. (25) but involving an admittance which is purely susceptive and such that the true function is purely real. This obviously reduces the computation required and leads to an important result discussed below. For this, the equation corresponding to eqn. (25), which involves the distribution of H_y in the plane $z = 0$, say G(y), is required. This equation is

$$\frac{1}{Y_A} = \frac{\frac{a}{2}\sum\limits_{s=1}^{\infty}\left[\int_0^a G(y')\phi_s(y')dy'\right]^2\Big/ Y_s + \sum\limits_{n=2}^{\infty}\left[\int_0^a G(y')\sin(n\pi y'/a)dy'\right]^2\Big/ Y_{0n}}{\left[\int_0^a G(y')\sin(\pi y'/a)dy'\right]^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

which again has the property that Y_A is stationary with respect to first-order variations of the function G(y) about its true value. For symmetrical junctions the values of the susceptance calculated by inserting approximations to the functions F(y) and G(y) in eqns. (25) and (27) respectively lie on different sides of the true value. The analysis in such cases therefore gives not only an approximate value for the junction susceptance but also the range within which the true value must lie. In the general case considered here Y_A is a complex quantity, and a comparison of eqns. (25) and (27) cannot give limits on the true value. It is still preferable to use both equations, however, since useful information can sometimes be obtained by observing the changes in Y_A as better approximations to the functions F(y) and G(y) are introduced into eqns. (25) and (27) respectively.

The calculation of Y_A from eqn. (25) and/or eqn. (27) is insufficient to complete the specification of the equivalent circuit, because it gives no indication of the value of the length v of the transmission line corresponding to the propagated mode in the partially filled waveguide. An obvious method by which v can be obtained is to repeat the analysis with a wave incident in the partially filled waveguide: this is an extravagant process, since it also gives the turns ratio n, already found by the first calculations.

The direct application of the variational technique to an asymmetrical waveguide junction therefore has the following three disadvantages:

(a) The unknown field functions are complex.

(b) The method fails to give unequivocal limits for the circuit parameters.

(c) The calculation of the third parameter is unnecessarily tedious.

A modification which removes the first two objections and minimizes the computational labour has been developed, and is the theoretical counterpart of the experimental procedure suggested by Weissfloch for the measurement of the equivalent-circuit parameters. This experimental procedure will therefore be described before the modified variational approach.

(4) EXPERIMENTAL DETERMINATION OF THE EQUIVALENT-CIRCUIT PARAMETERS[3]

The properties of a waveguide junction may be found experimentally by plotting the position of a minimum of the standing-wave pattern in one waveguide against the position of a short-circuit in the other waveguide. The theoretical shape of this curve is readily calculated in terms of the equivalent-circuit parameters: conversely, these parameters may be deduced from a curve plotted by experiment.

Suppose that the short-circuit is placed at $z = s$ in the waveguide structure shown in Fig. 2, and that a zero of the standing-wave pattern occurs at $z = -d$. The equivalent circuit of the junction is taken as in Fig. 3, the characteristic impedances of the transmission lines being made equal to the respective wave impedances for convenience,

i.e. $$Z_1 = 1/Y_{01}\colon\ Z_2 = 1/Y_1 \quad . \quad . \quad . \quad . \quad (28)$$

so that $$\frac{Z_2}{Z_1} = \frac{Y_{01}}{Y_1} = \frac{\beta_{01}}{\beta_1} \quad . \quad . \quad . \quad . \quad . \quad (29)$$

from eqns. (4) and (71).

When the short-circuit is placed at $z = s$ the impedance at the input to the transformer shown in Fig. 3 is

$$Z' = jn^2Z_2 \tan[\beta_1(s + v)] \quad . \quad . \quad . \quad (30)$$

and the condition that there should be a voltage zero at $z = -d$ is

$$Z' = -jZ_1 \tan[\beta_{01}(d + u)] \quad . \quad . \quad . \quad . \quad (31)$$

Eliminating Z' between these two equations gives

$$\tan[\beta_{01}(d + u)] = -\frac{n^2Z_2}{Z_1}\tan[\beta_1(s + v)] \quad . \quad (32)$$

which is the theoretical relation between d and s. Eqn. (32) may be rewritten

$$\tan(\theta + \theta_0) = -(n^2Z_2/Z_1)\tan(\phi + \phi_0) \quad . \quad . \quad (33)$$

where $$\theta = \beta_{01}d \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

$$\theta_0 = \beta_{01}u \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

$$\phi = \beta_1 s \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

$$\phi_0 = \beta_1 v \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

The curve of θ against ϕ oscillates about a straight line of slope -1: the amplitude of the oscillation depends on the ratio n^2Z_2/Z_1 and is small when this ratio is near unity. It is then more convenient to plot $(\theta + \phi)$ against θ, giving the type of curve shown in Fig. 4. The amplitude w, of the oscillation is given by

$$n[Z_2/Z_1]^{1/2} = \tan(\pi/4 - w/4) \quad . \quad . \quad . \quad (38)$$

If P is the point on the curve such that the slope has its maximum negative value, then

$$\phi_P = l\pi - \beta_1 v \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

and $$\theta_P = m\pi - \beta_{01}u \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

l and m being any integers.

The horizontal line through P lies midway between the maximum and minimum values of $(\theta + \phi)$.

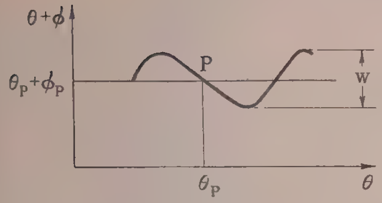


Fig. 4.—Typical curve showing the variation of $(\theta + \phi)$ against θ.

The circuit parameters n, u and v may be obtained by plotting he curve from experimental results, observing w, ϕ_P and θ_P and ien using eqns. (38)–(40). An alternative circuit which behaves ι exactly the same way as that above is given by

$$n' = nZ_1/Z_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

$$\beta_1 v' = \beta_1 v + \pi/2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

$$\beta_{01} u' = \beta_{01} u + \pi/2 \quad . \quad . \quad . \quad . \quad . \quad (43)$$

he particular circuit chosen is usually such as will make the ırns ratio as near unity as possible.

(5) MODIFIED VARIATIONAL METHOD

In Section 3 the final result is an expression for the input dmittance at the plane $z = 0$ (Fig. 2), when the partially filled aveguide extends indefinitely to the right of the Figure. The ıodified method consists of finding a similar expression for the ıput admittance when the partially filled waveguide is terminated y a short-circuit. This choice of termination immediately iminates the first objection listed in Section 3, for the structure ɔ the right of the plane $z = 0$ constitutes a lossless junction and y a general theorem the electric field is everywhere in phase ithin the junction.[7] The electric field may therefore be represented by a real function and a further result shows that the ıagnetic field is then expressed by an imaginary function. hese results are the field equivalents of the circuit condition hereby the voltage and current differ in phase by 90° for a ɜactive element.

Suppose, then, that the partially filled waveguide is terminated y a short-circuit placed at $z = s$. If s is sufficiently great, the vanescent modes excited by the junction are effectively zero at ıe position of the short-circuit. The propagated mode is ɛflected by the short-circuit, the phase of the reflected wave ɛing such that E_x must vanish for $z = s$. The fields in the partially filled waveguide now have the form [cf. eqns. (16) and (17)]

$$E_x = B_1'\phi_1(y)\sin\beta_1(s-z) + \sum_{n=2}^{\infty} B_n'\phi_n(y)\varepsilon^{-\alpha_n z} \quad . \quad (44)$$

$$H_y = -jB_1'Y_1\phi_1(y)\cos\beta_1(s-z) + \sum_{n=2}^{\infty} Y_n B_n'\phi_n(y)\varepsilon^{-\alpha_n z} \quad (45)$$

The fields in the empty waveguide have the same general form as in eqns. (14) and (15): since all the incident power is now totally reflected from the junction, the coefficient A_1 must have the same amplitude as A. The analysis in Section 3 may now be repeated and leads to an expression for the input admittance at $z = 0$ similar in form to eqn. (26). This admittance must be a pure susceptance, jB_A, since the junction is lossless. Furthermore, the distance of a minimum from the reference plane, say d, is given by

$$Y_{01}\cot(\beta_{01}d) = B_A \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

so that the equation corresponding to eqn. (25) can be written

$$\cot\beta_{01}d = -\frac{\left\{j\sum_{n=2}^{\infty} Y_{0n}\left[\int_0^a F(y')\sin(n\pi y'/a)dy'\right]^2 + \frac{ja}{2}\sum_{s=2}^{\infty} Y_s\left[\int_0^a F(y')\phi_s(y')dy'\right]^2 + \frac{a}{2}Y_1\cot\beta_1 s\left[\int_0^a F(y')\phi_1(y')dy'\right]^2\right\}}{Y_{01}\left[\int_0^a F(y')\sin(\pi y'/a)dy'\right]} \quad (47)$$

in which F(y) is again the electric field in the plane $z = 0$. If the magnetic field is taken as the starting-point, it is found [corresponding to eqn. (27)] that

$$\tan\beta_{01}d = -\frac{\left\{j\sum_{n=2}^{\infty}\left[\int_0^a F(y')\sin(n\pi y'/a)dy'\right]^2\Big/Y_{0n} + \frac{ja}{2}\sum_{s=2}^{\infty}\left[\int_0^a F(y')\phi_s(y')dy'\right]^2\Big/Y_s + \frac{a}{2Y_1}\tan\beta_1 s\left[\int_0^a F(y')\phi_1(y')dy'\right]^2\right\}}{\frac{1}{Y_{01}}\left[\int_0^a F(y'\sin(\pi y'/a)dy'\right]^2} \quad (48)$$

The expressions on the right-hand sides of eqns. (47) and (48) are again stationary with respect to variations of the field distributions about the true values. A second differentiation shows that these expressions are minima for the true values, so that the approximations to cot $\beta_{01}d$ calculated from eqns. (47) and (48) will be, respectively, too large and too small. The true value must lie somewhere between the approximations calculated from the two equations, so that at any stage the maximum error due to the approximation can be readily found.

From the above analysis a theoretical curve showing the dependence of the minimum position on the short-circuit position can be calculated. The parameters of the equivalent circuit can then be found as described in Section 4. Since the maximum error in the curve is known, the corresponding maximum error in the circuit parameters can be calculated. The second of the objections to the original method is thus also removed by this modified approach.

A variational expression has to be evaluated for each point of the curve of $(d + s)$ against s, so that it might be thought that the computational labour would be greater than for the original method. However, much of the computation, e.g. the evaluation of the integrals, need be done only once, and this coupled with the need to handle only real quantities means that there is little, if any, more work required.

(6) NUMERICAL EXAMPLE

The equivalent circuit has been derived for the junction between an empty waveguide of internal cross-section 1 × 0·5 in and the same guide containing a slab of dielectric, 0·7 cm thick and of relative permittivity 2·47, positioned as shown in Fig. 2. The results have been obtained for a free-space wavelength of 3·14 cm. The evaluation of the variational expressions has been carried through in the manner described by Schwinger for symmetrical junctions:[4] the functions F(y) and G(y) have each been approximated by an expression of the type $\sin(\pi y/a) + p \sin(2\pi y/a)$, p being selected in each case to give the required stationary property. Numerical results for $\beta_{01}d$ for selected values of $\cot \beta_1 s$ are shown in Table 1. The maximum error resulting from the approximation used is $\pm 0\cdot03$ rad in $\beta_{01}d$, i.e. $\pm 0\cdot02$ cm in d.

Table 1

NUMERICAL VALUES FOR $\beta_{01}d$

[Average values obtained from eqns. (42) and (43)]

$\cot \beta_1 s$	−4·0	−2·0	−1·0	−0·5	−0·2	0·2	0·5	1·0	2·0	4·0
$\beta_{01}d$	3·242	3·459	3·801	4·190	4·519	1·832	2·103	2·391	2·656	2·844
Possible error in $\beta_{01}d$	±0·018	±0·026	±0·030	±0·025	±0·018	±0·007	±0·002	±0·004	±0·009	±0·021
$\beta_{01}d + \beta_1 s$	−0·145	−0·147	−0·127	−0·059	+0·003	0·064	0·068	0·035	−0·022	−0·053

The curve of $\beta_1 s + \beta_{01}d$ against $\beta_{01}d$ obtained from these points is shown in Fig. 5, from which it follows that

$$\left.\begin{aligned} w &= 0\cdot23 \pm 0\cdot03 \\ \theta_p &= 2\cdot72 \pm 0\cdot02 \\ \phi_p &= -2\cdot76 \pm 0\cdot02 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (49)$$

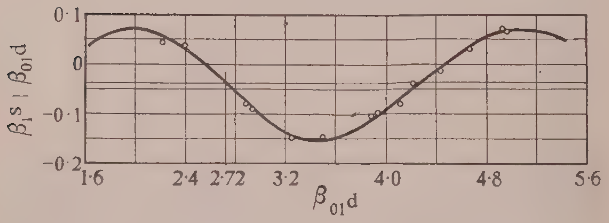


Fig. 5.—Calculated curve of $(\beta_1 s + \beta_{01}d)$ against $\beta_{01}d$.
○○○ Experimental points.

The equivalent-circuit parameters as deduced from eqns. (38)–(40) are therefore

$$n[Z_2/Z_1]^{1/2} = 0\cdot89 \pm 0\cdot01 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

$$\beta_{01}u = 0\cdot42 \pm 0\cdot02 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (51)$$

$$\beta_1 v = 2\cdot76 \pm 0\cdot02 \text{ or } -0\cdot38 \pm 0\cdot02 \quad . \quad (52)$$

the limits being estimated directly from the limits in Table 1.

When the characteristic impedances of the transmission lines are made equal to the corresponding wave impedances, eqn. (29) applies and gives, for the present example,

$$Z_2/Z_1 = 0\cdot81 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

From eqns. (50) and (53) n is found to be $0\cdot99 \pm 0\cdot01$, i.e. the turns ratio of the transformer in the equivalent circuit may be taken as unity within the limits of accuracy involved in the present calculation. The reflection coefficient therefore has the magnitude predicted by inserting wave impedances into the ordinary transmission-line formula.

There is an uncertainty of 180° in the phase of the transmitted wave because of the multiples of π in eqns. (39) and (40). This uncertainty arises because the calculation is based on impedances, which are invariant when transformed through a $\lambda/2$ section of transmission line. In an equivalent circuit of the type shown in Fig. 1(b), the connections to either side of the transformer may be reversed without altering the conditions on the input side, but causing a change of 180° in the phase of the wave on the output side. The uncertainty may in practice be resolved by stipulating that the transformer is so connected that there is no phase difference between its primary and secondary voltages and then selecting the multiples of π associated with eqns. (39) and (40) to give the most appropriate total phase-change across the junction. In the present example the two alternatives are typified by the choice of values available for v in eqn. (52). The first of these implies a total phase-change of 3·18 rad across the junction plane and the second a total phase-change of 0·04 rad. The latter is physically much the more likely so that the values for u and v are $0\cdot067\lambda_{01}$ and $-0\cdot061\lambda_1$ respectively: λ_{01} and λ_1 are the wavelengths in the empty and partially filled waveguides respectively.

As a check on the theoretical calculations, measurements were made on a similar structure. The curve of $(d + s)$ against s was obtained by inserting a piece of dielectric of the appropriate cross-section in a short-circuited waveguide. The standing-wave minimum position was observed for a succession of different lengths of the dielectric, obtained by machining successive pieces off the original slab. The experimental points are shown in Fig. 5 and do not depart from the theoretical curve by more than 0·02 rad: this is within the estimated accuracy of the apparatus used.

The fact that the reflection coefficient may be accurately predicted by inserting wave impedances in the ordinary transmission-line formula is of great value in designing matching transformers using waveguides partially filled with dielectric. Although there is no obvious reason why this result should be exact, numerical calculations for other similar junctions have confirmed its validity to within one or two per cent provided that both waveguides have the same cross-sectional dimensions. The following qualitative discussion suggests that the results may be generally valid for junctions of the type discussed in the paper. When an empty waveguide is joined to one of the same dimensions completely filled with a homogeneous material, the transverse field patterns for the dominant modes are identical in the two waveguides. The boundary conditions at the interface can be satisfied by the dominant modes alone, so that no additional reactive fields are established; the equivalent circuit consists of a direct connection between the two transmission lines corresponding to the dominant modes, so that the reflection coefficient is given actly by substituting wave impedances in the transmission-line flection formula. If, however, one of the waveguides is not mpletely filled, the transverse pattern for the dominant mode ffers from that in an empty guide and, as was seen earlier in e paper, additional evanescent fields are established. Since e discontinuity occurs only in a plane transverse to the direction propagation, it is plausible to suppose from the well-known operties of irises that the major effect of the evanescent fields represented by a shunt susceptance located at the junction tween the transmission lines corresponding to the dominant aveguide modes. In other words, in the circuit shown in Fig. c) the most important component is the shunt one, X_3, so that $_1$ and X_2 may be assumed zero as a reasonably good first pproximation. Furthermore, if the discontinuity is not a great ne, the ratio Z_1/X_3 is small compared with unity. A straightrward calculation shows that the phase changes resulting from e introduction of X_3 are proportional to Z_1/X_3 while the change the magnitude of the reflection coefficient is proportional to $_1/X_3)^2$. This means that the magnitude of the reflection efficient is given correctly to the first order of magnitude by the ansmission-line formula using wave impedances.

(7) GENERAL DISCUSSION

The modified variational method which has been described vercomes the disadvantages of a straightforward approach to e calculation of the equivalent-circuit parameters of an axially nsymmetrical junction. A relatively crude approximation to e aperture fields leads to quite accurate values of the parameters s illustrated by the example considered here. The method can e applied to any axially unsymmetrical junction, the only ossible further complication being that, in general, the fields will epend on both the transverse co-ordinates, x and y.

(8) ACKNOWLEDGMENT

One of the authors, Dr. R. E. Collin, is indebted to the Athlone ommittee for a Fellowship, which was held during the period f the work described.

(9) REFERENCES

) MONTGOMERY, C. G., DICKE, R. H., and PURCELL, E. M.: "Principles of Microwave Circuits" (Radiation Laboratory Series, McGraw-Hill Book Company, 1948), Vol. 8, Chapter 5.

2) MARCUVITZ, N.: "Waveguide Handbook" (Radiation Laboratory Series, McGraw-Hill Book Company, 1951), Vol. 10, Section 3.3.

3) *Ibid.*, Section 3.4.

4) SCHWINGER, J.: "Discontinuities in Waveguides" (M.I.T. Lecture Notes edited by D. S. Saxon).

5) MILES, J. W.: "The Equivalent Circuit of a Plane Discontinuity in a Cylindrical Waveguide," *Proceedings of the Institute of Radio Engineers*, 1946, **34**, p. 728.

6) LEWIN, L.: "Advanced Waveguide Theory" (Iliffe, London, 1950).

7) MONTGOMERY, G. G., *et al.*: *ibid.*, Section 5.5.

(10) APPENDICES

0.1) Possible Modes within a Waveguide Partially Filled with Dielectric

The modes of propagation within a waveguide partially filled ith dielectric have been studied by a number of authors. In e present analysis only fields which are independent of x and aving the components E_x, H_y and H_z need be considered. Maxwell's electromagnetic equations then simplify to

$$j\omega\mu H_y = -\partial E_x/\partial z \quad . \quad . \quad . \quad . \quad . \quad (53)$$

$$j\omega\mu H_z = +\partial E_x/\partial y \quad . \quad . \quad . \quad . \quad . \quad (54)$$

together with

$$\frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} = \beta_0^2 E_x \text{ for } 0 \leqslant y \leqslant a - t \quad . \quad . \quad (55)$$

$$\frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} = \epsilon_r \beta_0^2 E_x \text{ for } a - t \leqslant y \leqslant a \quad . \quad (56)$$

where ϵ_r is the relative permittivity of the dielectric slab, which fills the region of the waveguide cross-section defined by

$$\text{and} \quad \left.\begin{array}{c} 0 \leqslant x \leqslant b \\ a - t \leqslant y \leqslant a \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (57)$$

The solution must also satisfy the boundary conditions that E_x vanishes when y is equal to 0 or a, and that E_x and H_z are continuous at the boundary $y = (a - t)$ between the empty and filled portions of the waveguide.

In solutions representing waves propagating or evanescent in the direction of the z-axis, the fields depend on z only through a factor of the type, $\varepsilon^{-\gamma z}$, γ being a propagation coefficient. The field in the empty portion of the guide is then of the form

$$E_x = C \sin [(\beta_0^2 + \gamma^2)^{1/2} y] \varepsilon^{-\gamma z} \quad (0 \leqslant y \leqslant a - t) \quad . \quad (58)$$

which satisfies eqn. (55) and the boundary condition at $y = 0$. Similarly, in the filled region of the waveguide

$$E_x = D \sin [(\epsilon_r \beta_0^2 + \gamma^2)^{1/2} (a - y)] \varepsilon^{-\gamma z} \quad (a - t \leqslant y \leqslant a) \quad (59)$$

which satisfies eqn. (56) and the boundary condition at $y = a$.

From eqns. (54), (58) and (59), the component H_z is given by

$$j\omega\mu H_z = +(\beta_0^2 + \gamma_2^2)^{1/2} C \cos [(\beta_0 + \gamma^2)^{1/2} y] \varepsilon^{-\gamma z} \qquad (0 \leqslant y \leqslant a - t) \quad . \quad (60)$$

$$j\omega\mu H_z = -(\epsilon_r \beta_0^2 + \gamma^2)^{1/2} D \cos [(\epsilon_r \beta_0^2 + \gamma^2)^{1/2} (a - y)] \varepsilon^{-z\gamma} \qquad (a - t \leqslant y \leqslant a) \quad . \quad (61)$$

The components E_x and H_z are continuous at $y = (a - t)$, so that

$$C \sin [(\beta_0^2 + \gamma^2)^{1/2} (a - t)] = D \sin [(\epsilon_r \beta_0^2 + \gamma^2)^{1/2} t] \qquad (62)$$

and

$$(\beta_0^2 + \gamma^2)^{1/2} C \cos [(\beta_0^2 + \gamma^2)^{1/2} (a - t)] = -(\epsilon_r \beta_0^2 + \gamma^2)^{1/2} D \cos [(\epsilon_r \beta_0^2 + \gamma^2)^{1/2} t] \quad . \quad (63)$$

From these two equations,

$$(\epsilon_r \beta_0^2 + \gamma^2)^{1/2} \tan [(\beta_0^2 + \gamma^2)(a - t)] = -(\beta_0^2 + \gamma^2)^{1/2} \tan [(\epsilon_r \beta_0^2 + \gamma^2)^{1/2} t] \quad . \quad (64)$$

This equation is satisfied by an infinite set of values of γ^2; if ϵ_r and t are not too large, one of these values is negative, so that γ may be written $j\beta_1$, i.e. the phase coefficient of the propagated mode. The remaining values of γ^2 are positive and may be arranged in ascending order giving $\alpha_2^2, \alpha_3^2 \ldots$ etc., the attenuation coefficients of the evanescent modes.

Once the propagation constant γ has been determined, the ratio D/C may be calculated from eqn. (62) and then the component E_x may be written

$$E_x = \phi_n(y) \varepsilon^{-\gamma z} \quad . \quad . \quad . \quad . \quad . \quad . \quad (65)$$

where

$$\phi_n(y) = C_n \sin[(\beta_0^2 + \gamma^2)^{1/2}y] \quad (0 \leqslant y \leqslant a - t) \quad . \quad (66)$$

$$= C_n \frac{\sin[(\beta_0^2 + \gamma^2)^{1/2}(a - t)]}{\sin[(\epsilon_r\beta_0^2 + \gamma^2)^{1/2}t]} \sin[(\epsilon_r\beta_0^2 + \gamma^2)^{1/2}(a - y)]$$

$$(a - t \leqslant y \leqslant a) \quad . \quad (67)$$

in which γ is to be replaced by $j\beta_1$ if $n = 1$ and by α_n if n is 2, 3 . . . etc. The constant C_n may be selected to satisfy the condition that

$$\int_0^a \phi_n^2(y)dy = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (68)$$

The result

$$\int_0^a \phi_m(y)\phi_n(y)dy = 0 \text{ if } m \neq n \quad . \quad . \quad . \quad (70)$$

is a direct consequence of a general theorem on the orthogonality of the field distributions of waveguide modes and may also be verified by direct integration coupled with eqn. (65) from which the propagation coefficients are calculated. The set of functions of y is complete in the sense that any arbitrary function of y may be expanded in a series of the type $\sum_{n=1}^{\infty} c_n\phi_n(y)$. From eqns. (53) and (65),

$$H_y = \gamma E_x/j\omega\mu = -j\gamma Y_0 E_x/\beta_0 \quad . \quad . \quad . \quad . \quad (71)$$

The wave admittances are therefore

$$Y_1 = \beta_1 Y_0/\beta_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (72)$$

$$Y_n = -j\alpha_n Y_0/\beta_0 \quad (n = 2, 3 \ldots) \quad . \quad . \quad (73)$$

(10.2) The Stationary Property of the Expression for the Input Admittance

The input admittance is given by eqn. (26)

$$Y_A = \frac{\frac{1}{2}a \sum_{s=1}^{\infty} Y_s \left[\int_0^a F(y')\phi_s(y')dy'\right]^2 + \sum_{n=2}^{\infty} Y_{0n}\left[\int_0^a F(y')\sin(n\pi y'/a)dy'\right]^2}{\left[\int_0^a F(y')\sin(\pi y'/a)dy'\right]^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (74)$$

If the function $F(y')$ changes by the small amount $\delta F(y')$, then Y_A as calculated from this equation will change by an amount δY_A, where

$$\delta Y_A \left[\int_0^a F(y')\sin(\pi y'/a)dy'\right]^2$$
$$+ 2Y_A\left[\int_0^a F(y')\sin(\pi y'/a)dy'\right]\left[\int_0^a \delta F(y')\sin(\pi y'/a)dy'\right]$$
$$= a\sum_{s=1}^{\infty} Y_s\left[\int_0^a F(y')\phi_s(y')dy'\right]\left[\int_0^a \delta F(y')\phi_s(y')dy'\right]$$
$$+ 2\sum_{n=2}^{\infty} Y_{0n}\left[\int_0^a F(y')\sin(n\pi y'/a)dy'\right]$$
$$\left[\int_0^a \delta F(y')\sin(n\pi y'/a)dy'\right] \quad . \quad (75)$$

Replacing Y_A and $\int_0^a F(y')\sin(\pi y'/a)dy'$ by their expressions in terms of A and A_1 and rearranging the order of the terms gives

$$\delta Y_A\left[\int_0^a F(y')\sin(\pi y'/a)dy'\right]^2$$
$$= -\int_0^a \delta F(y)\Bigg[a(A - A_1)Y_{01}\sin(\pi y/a)$$
$$- 2\sum_{n=2}^{\infty} Y_{0n}\int_0^a F(y')\sin(n\pi y'/a)\sin(n\pi y/a)dy'$$
$$- a\sum_{s=1}^{\infty} Y_s\int_0^a F(y')\phi_s(y')\phi_s(y)dy'\Bigg]dy \quad . \quad . \quad . \quad . \quad (76)$$

from which it follows that the right-hand side vanishes if F(y) satisfies the integral eqn. (23). The input admittance Y_A as calculated from eqn. (74) is therefore stationary, i.e. $\delta Y_A = 0$, with respect to variations of the function F(y) about its true value.

621.373.42: 621.316.729: 621.38

Monograph No. 146 R
Aug. 1955

DISCRIMINATION OF A SYNCHRONIZED OSCILLATOR AGAINST INTERFERING TONES AND NOISE

By D. G. TUCKER, D.Sc., Member, and G. G. JAMIESON.

(*The paper was first received* 29*th December*, 1954, *and in revised form* 13*th May*, 1955. *It was published as an* Institution Monograph *in August*, 1955.)

SUMMARY

An account is given of the discrimination provided by a synchronized oscillator against unwanted signals accompanying the synchronizing tone; the synchronized oscillator is more specifically a non-linear regenerative tuned circuit. The discrimination is partly due to the frequency response of the system, but has also an important contribution from the non-linear behaviour of the circuit, provided that the wanted (sychronizing) signal has a greater amplitude, after allowing for frequency response, than the unwanted (interfering) signals.

Section 2 gives a general description of the phenomena of synchronization and non-linear discrimination, illustrated by experimental observations of frequency response and non-linear discrimination in resistance-capacitance-tuned oscillators. Both single-frequency and noise interferences are considered. It is concluded that very large amounts of discrimination can be obtained when the natural frequency of the oscillator is very close to the synchronizing frequency, so that only very small amplitudes of the latter are needed to maintain synchronism. In more practical cases, where reasonable amounts of variation of both natural and synchronizing frequencies must be allowed for, only smaller amounts of discrimination due to non-linearity can be obtained, but 10 dB or more is quite feasible. This amount will be important when the interfering frequencies are so close to the synchronized frequency that the amount of frequency discrimination is negligible. For instance, use is made of these effects in the homodyne and synchrodyne demodulators. It is shown that when the synchronizing tone is not dominant or is absent no suppression of the interference or narrowing of its spectrum is obtained.

Section 3 gives a mathematical analysis of the effects, on the assumption that the non-linear law of the system has no terms above the cubic, and that the feedback loop has uniform amplitude and phase response over the frequency band covered by the applied signals. Equations are developed for the improvement of signal/noise ratio and suppression of single-tone interference. Interference due to amplitude modulation and phase modulation of the synchronizing tone is also dealt with, and it is shown that the suppression of amplitude modulation is greater than that of signals merely accompanying the synchronizing tone, but that phase modulation is not suppressed at all.

(1) INTRODUCTION

The synchronization of a valve oscillator to an external frequency not very different from its own natural frequency is a well-known phenomenon, and the simplest way of achieving it is by the injection of a suitable amplitude of the external frequency into the oscillating circuit. Many papers have been published describing and analysing this process: the first and most important was by Appleton;[1] a recent one by Gillies[2] gives a clear physical picture of most of the phenomena; and one by Tucker[3] gives a rather simpler analytical approach to the quantitative relationships of the system. The important conclusion from all the work is that the process of synchronization is a non-linear one; in very simple terms it can be said that the injection of a tone, at a frequency near the natural frequency of the highly-regenerative circuit, builds up a forced oscillation at an amplitude large enough to overload the valve and thereby to reduce its gain, so that the free oscillation is unable to continue. Then only the forced oscillation remains, and the oscillator is said to be synchronized. The synchronized oscillator is clearly a non-linear regenerative tuned circuit with the regeneration so large that free oscillation can occur in the absence of the synchronizing signal.

An important property of this circuit is that if other signals are mixed with the synchronizing tone, and even if they are very close to it in frequency, they can under certain circumstances be subjected by the circuit to a high degree of discrimination; i.e. the output contains relatively much less of these other signals than does the input. This discrimination can include the effect of the frequency response* of the circuit, but its most interesting and important feature is that it also includes a non-linear effect which gives discrimination against an unwanted signal even when the frequency response is flat. This latter property seems to have attracted little attention, there being only three papers[4,5,6] dealing with it as far as is known; yet the property is vital to demodulators of the homodyne and synchrodyne type. The papers referred to, however, deal only with the discrimination against single interfering tones or simple envelope-modulation components. For more complex interfering signals, such as noise, the treatment is more difficult, and in the limit it may be a question of what output is obtained when the input signal comprises noise alone without any coherent tone.

It should be noted that Reference 4 gives experimental results only for the discrimination against interference which modulates the synchronizing tone. The present paper gives experimental measurements only for interference accompanying the synchronizing tone, but the theoretical part deals with both cases; the paper is the account of an experimental and theoretical investigation into these various aspects of the response of a non-linear regenerative circuit to complex signals and noise. Section 2 gives first some brief notes on the basic relationships in synchronized oscillators, then some results of discrimination against pure tones, and finally measurements of discrimination against noise. Section 3 gives the theoretical analysis.

(2) GENERAL ACCOUNT OF NON-LINEAR PHENOMENA, AND EXPERIMENTAL RESULTS

(2.1) Synchronization and Forced Oscillations

A typical oscillator circuit is shown in Fig. 1. The amplitude of the synchronizing signal to be injected into the terminals shown in order to suppress the free oscillation completely will depend on the difference in frequency between the synchronizing signal and the free oscillation, and on the amplitude of the latter. If this injected amplitude is measured against frequency difference, a normalized curve, unique for each oscillator, may be plotted, giving a relationship similar to those of the typical measured

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Tucker was formerly, and Mr. Jamieson is, in the Royal Naval Scientific Service.

* The distinction between frequency response and discrimination should be noted. Frequency response is the variation of output as the frequency of a single input tone is varied; discrimination is the reduction of the level of one signal relative to another when both are simultaneously present.

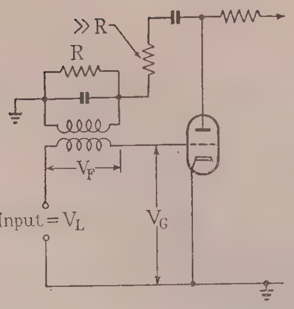


Fig. 1.—Synchronized oscillator circuit.

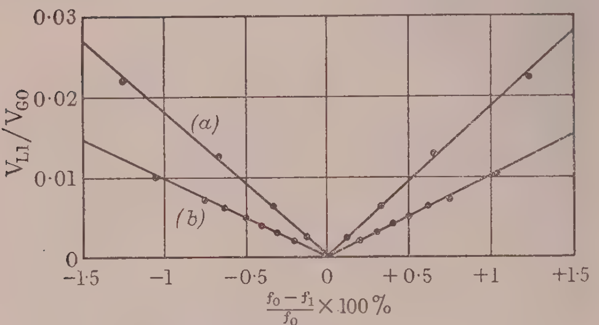


Fig. 2.—Synchronizing characteristics.

(*a*) For *RC* oscillator.
(*b*) For Wien-bridge oscillator.

curves of Fig. 2. In this Figure the injected amplitude, V_{L1}, has been expressed as a fraction of the natural oscillation voltage, V_{G0}, at that part of the circuit where the signal is injected, i.e. at the grid. An approximate curve may be calculated[3,5] from the equation

$$\frac{V_{L1}}{V_{G0}} = 2Q\frac{f_1 - f_0}{f_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where Q is the value appropriate to the frequency-selective element in the feedback path, f_0 is the natural frequency, and f_1 is the synchronizing frequency.

The free and forced oscillations may exist simultaneously in the circuit until V_{L1}/V_{G0} reaches the critical ratio given in eqn. (1); then the free oscillation is entirely suppressed, and the oscillator is said to be synchronized. The suppression of the free oscillation has been shown to take two forms:[2,3]

(*a*) For small frequency differences between forced and free oscillations, which require only small values of V_{L1}/V_{G0}, the process of synchronization can be considered as that of progressively changing the frequency of free oscillations with increase of V_{L1}/V_{G0} until the critical ratio is reached, when the free oscillation frequency coincides with the injected frequency. This effect is shown in Fig. 3, where experimental results are given.

(*b*) For larger frequency differences, and hence increased values of V_{L1}/V_{G0}, the amplitude of the free oscillation is reduced to zero before its frequency nears that of the synchronizing signal.

When the oscillator is in the synchronized condition the output of the forced oscillation will not, in general, have the same amplitude as the free oscillation it has replaced. The precise determination of this output characteristic is difficult, as it depends on the extent to which harmonics are produced in the circuit. The harmonics affect the shape and symmetry of the curve of output amplitude against frequency difference between forced and free oscillations. It can be taken as a guide that, if the value of V_{L1}/V_{G0} is kept adjusted to the minimum necessary for synchronization, the output of the forced oscillation will decrease as the frequency difference is increased.

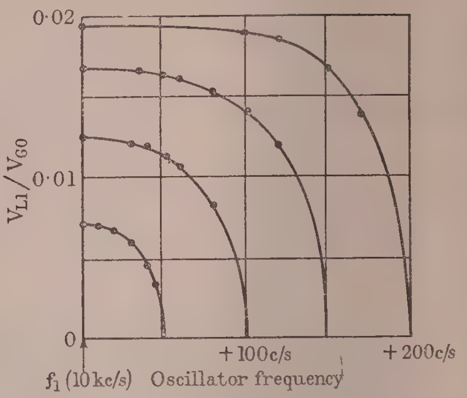


Fig. 3.—Change of oscillator frequency before synchronization (*RC* oscillator).

N.B. f_1 is the synchronizing frequency.

If the synchronizing frequency is different from that of free oscillation, there will be a phase difference between the input and the forced oscillation, i.e. between V_{L1} and V_{G1}. The phase angle is given approximately[3,5] by

$$\sin\theta = \frac{f_1 - f_0}{f' - f_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where f_0 is the free oscillation frequency, f_1 the injected frequency, and f' the frequency at which "pull out" from synchronization occurs.

Two criteria have been given[3] for the forced oscillation to remain stable, i.e. to completely suppress the free oscillation, namely

(*a*) The phase difference between V_{L1} and V_{G1} must be less than $\pi/2$.

(*b*) The grid voltage V_{G1} for the forced oscillation must be greater than $1/\sqrt{2}$ times the value (V_{G0}) when the free oscillation only is present.

The limits of the frequency range over which the oscillator will remain synchronized will be determined by the phase criterion for small values of V_{L1}/V_{G0}, the angle reaching $\pi/2$ before V_{G1} has decreased to the critical amplitude. For large values of V_{L1}/V_{G0} the quantity V_{G1} will reach the minimum value for stability before the phase angle nears $\pi/2$.

(2.2) Discrimination in the Circuit against Interference from a Single Tone accompanying the Synchronizing Signal

The use of a linear filter to provide discrimination between tones very close in frequency is particularly difficult if the frequency of the wanted signal is not precisely known or is subject to variation. In these circumstances the use of a synchronized oscillator can be attractive, as it provides a highly selective filter which uses a property of the circuit other than its frequency selectivity.

If it is assumed that the frequency-selective element in the feedback path has a flat amplitude response and zero phase-shift over the frequency band, and that the non-linear law of the limiting circuit can be represented by a cubic equation, the circuit can be analysed as shown in Reference (6). The main results of the paper cited are included in the more general analysis in

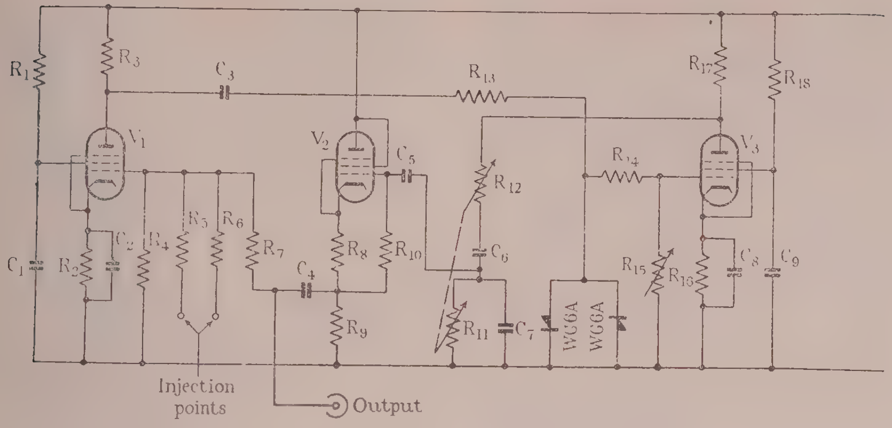


Fig. 4.—Circuit arrangement of Wien-bridge oscillator.

$R_1 = 22\,\text{k}\Omega$
$R_2 = 470\,\Omega$
$R_3 = 33\,\text{k}\Omega$
$R_4 = 8{\cdot}2\,\text{k}\Omega$
$R_5 = R_6 = R_7 = 100\,\text{k}\Omega$
$R_8 = 100\,\Omega$
$R_9 = 12\,\text{k}\Omega$

$R_{10} = R_{11} = R_{12} = 100\,\text{k}\Omega$
$R_{13} = 33\,\text{k}\Omega$
$R_{14} = 3{\cdot}3\,\text{k}\Omega$
$R_{15} = 1\,\text{k}\Omega$
$R_{16} = 220\,\Omega$
$R_{17} = 22\,\text{k}\Omega$
$R_{18} = 10\,\text{k}\Omega$

$C_1 = 0{\cdot}1\,\mu\text{F}$
$C_2 = 0{\cdot}5\,\mu\text{F}$
$C_3 = C_4 = C_5 = 0{\cdot}1\,\mu\text{F}$
$C_6 = C_7 = 220\,\mu\mu\text{F}$
$C_8 = 0{\cdot}5\,\mu\text{F}$
$C_9 = 0{\cdot}1\,\mu\text{F}$

$V_1 = \text{CV}329$
$V_2 = \text{CV}2127$
$V_3 = \text{CV}138$

Section 3 of the present paper. When the interference is caused by a single tone, the suppression of this tone between the input and the grid of the valve, relative to the synchronizing signal, is shown from eqn. (19) to be

$$\frac{1 - a_1 - \frac{3}{2}a_3V_{G1}^2}{1 - a_1 - \frac{3}{4}a_3V_{G1}^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where a_1 and a_3 are the linear and cubic coefficients respectively of the power series representing the loop gain, and V_{G1} is the voltage of the synchronizing frequency at the grid. This can be expressed alternatively as

$$2 + (a_1 - 1)\frac{V_{G1}}{V_{L1}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where V_{L1} is the injected voltage of synchronizing signal.

This expression clearly shows that when the loop gain is unity ($a_1 = 1$) the suppression is 6 dB. With $a_1 > 1$, i.e. with the circuit capable of freely oscillating in the absence of synchronization, large values of suppression can be obtained if V_{L1}/V_{G0} (which is almost the same as V_{L1}/V_{G1}) is kept very small. In practice, V_{L1}/V_{G0} cannot be made indefinitely small, since the amplitude of the synchronizing signal injected must be sufficient to keep the oscillator synchronized over the maximum frequency difference expected between the synchronizing signal and the free oscillation. Even if this difference is nominally zero, considerations of drift, etc., in the circuit used will fix a minimum value for this injection ratio.

The simple explanation of the cause of this large discrimination is this: (*a*) When two signals are applied to a non-linear circuit, e.g. a cube-law circuit, the gain provided to any one signal is diminished by a factor proportional to the square of the amplitude of that signal plus twice the square of the other signal. Thus if one signal is of a much lower amplitude than the other, its gain is diminished nearly twice as much as that of the other. (*b*) This difference in gain is then multiplied, perhaps very greatly, by the effect of regeneration in a manner analogous to that by which frequency response is multiplied.

In order to minimize the effects of frequency selectivity so that the non-linear phenomena could be isolated, the oscillators used for practical measurements were of the resistance-capacitance-tuned type, one with a Wien bridge as the frequency-dependent element in the feedback path, and the other with a 3-stage *RC* network. Both these frequency-selective networks have very low effective Q-factors (obtained by consideration of the slope of the phase characteristic in the region of the oscillation frequency). The practical circuits are shown in Figs. 4 and 5 respectively.

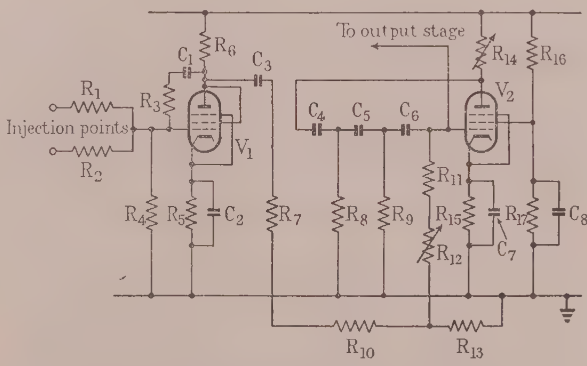


Fig. 5.—Circuit arrangement of *RC* oscillator.

$R_1 = R_2 = R_3 = 100\,\text{k}\Omega$
$R_4 = 1\,\text{M}\Omega$
$R_5 = 470\,\Omega$
$R_6 = 100\,\text{k}\Omega$
$R_7 = 56\,\text{k}\Omega$
$R_8 = R_9 = 120\,\text{k}\Omega$
$R_{10} = 470\,\Omega$
$R_{11} = 120\,\text{k}\Omega$
$R_{12} = 5\,\text{k}\Omega$
$R_{13} = 68\,\Omega$
$R_{14} = 30\,\text{k}\Omega$

$R_{15} = 330\,\Omega$
$R_{16} = R_{12} = 10\,\text{k}\Omega$

$C_1 = 0{\cdot}1\,\mu\text{F}$
$C_2 = 1\,\mu\text{F}$
$C_3 = 0{\cdot}1\,\mu\text{F}$
$C_4 = C_5 = C_6 = 47\,\mu\mu\text{F}$
$C_7 = 0{\cdot}5\,\mu\text{F}$
$C_8 = 0{\cdot}1\,\mu\text{F}$

$V_1 = V_2 = \text{CV}138$

The Wien-bridge oscillator was amplitude-limited by the non-linear law of two germanium rectifiers, while in the *RC* oscillator* the curve of the pentode valve characteristic was used, so that if, as seems likely, the relationships between the coefficients in the

* Both oscillators are, of course, of the resistance-capacitance type. The names "Wien bridge" and "*RC*" are used to distinguish them.

non-linear laws are not the same, different suppression characteristics can be expected.

It has been assumed for the purpose of the mathematical analysis in Section 3 that the effect of frequency selectivity in the oscillator circuit is negligible, but it is evident from the experimental results for the Wien-bridge oscillator given in Fig. 6 that, although the frequency selectivity of the feedback network itself is very low, the overall frequency selectivity has been considerably enhanced by the positive feedback, as shown by the departure of the curves from the horizontal. This means that the non-linear suppression cannot be completely isolated, but it may be inferred that the discrimination due to the non-linearity only is given by the intercept of the curves on the axis at zero frequency-difference. It will be seen that the curves for different values of V_{L1}/V_{G0} run almost parallel, which indicates that the frequency discrimination appears to be merely a constant addition to the non-linear discrimination, the existence of the latter being proved by the dependence of the amount of suppression on V_{L1}. This is also illustrated in Fig. 7, which shows the measured suppression against V_{L1}/V_{G0} for the Wien-bridge oscillator at two values of frequency difference between the synchronizing tone, f_1, and the unwanted frequency, f_2. The curves are nearly parallel over the useful range. The intercept of the curves on the axis at zero frequency-difference in Fig. 6 have been plotted in Fig. 7, and this curve will represent that portion of the total suppression due to the non-linearity.

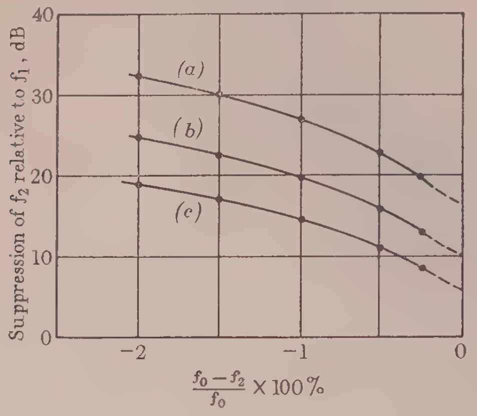


Fig. 6.—Effect of frequency response on the suppression of the unwanted tone.

Wien-bridge oscillator, $f_1 = f_0 = 10$ kc/s.
$V_{L2}/V_{L1} = 0{\cdot}1$.
(a) $V_{L1}/V_{G0} = 0{\cdot}000\,75$
(b) $V_{L1}/V_{G0} = 0{\cdot}001\,85$
(c) $V_{L1}/V_{G0} = 0{\cdot}003\,8$

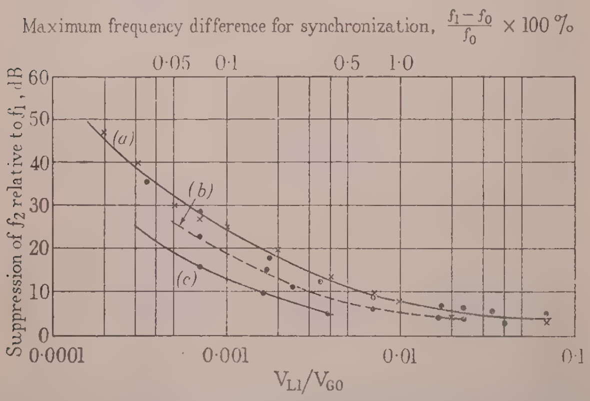


Fig. 7.—Suppression of unwanted tone.

Wien-bridge oscillator.
(a) $f_1 - f_2 = 100$ c/s.
● is measured point for $V_{G0} = 290$ mV
× is measured point for $V_{G0} = 1$ volt
(b) $f_1 - f_2 = 50$ c/s.
(c) Non-linear discrimination.

Table 1 shows that the effect of detuning the natural frequency of the oscillator from the synchronizing frequency is to limit the maximum suppression obtainable approximately to that value given for the ratio V_{L1}/V_{G0} necessary to effect synchronization. By reference to Fig. 2, a scale of frequency difference between f_1 and f_0 as the abscissa can be substituted in Fig. 7, as shown.

Table 1

Suppression of Unwanted Signal in Relation to Frequency Difference between Natural and Synchronized Oscillations (Wien-Bridge Oscillator) $V_{L2}/V_{L1} = 0{\cdot}1$ $(f_1 - f_2) = 100$ c/s

		$f_1 = f_0$		$f_0 - f_1 = 10$ c/s		$f_0 - f_1 = 30$ c/s		$f_0 - f_1 = 70$ c/s	
V_{L1}	V_{G1}	V_{G2}	Suppression	V_{G2}	Suppression	V_{G2}	Suppression	V_{G2}	Suppression
mV	volts	mV	dB	mV	dB	mV	dB	mV	dB
0·1	1								
0·2		0·5	46						
0·3		1·3	37·8						
0·5		2·5	32						
0·7		4·2	27·5						
1·0		7·0	23·1						
1·6				10	20				
2·0		15·0	16·5	15	16·5				
3·7						25	12		
4·0	1	28·0	11	30	10·4	28	11		
7·0		43	7·3	47	6·6	44	7·1		
7·5								59	4·6
10·0		55	5·2	54	5·3	53	5·5	67	3·5
20	1·02	68	3·3	68	3·3	68	3·3	72	2·9
40	1·03	73	2·7	73	2·7	73	2·7	74	2·6
70	1·05	75	2·5	74	2·6	76	2·4	78	2·1
100	1·07	84	1·5	82	1·7	83	1·5	82	1·7

It should be noted that when detuning of the oscillator can occur the frequency of the interfering tone should preferably be outside the permitted range of natural frequency. If this condition is not regarded, it will be possible for the natural frequency to coincide with the interfering frequency. When this happens, the interfering frequency is more favourably placed than the synchronizing frequency and may be able to cause discrimination against the wanted signal. Whether this can occur or not depends on the relative levels of the interfering and synchronizing tones. If the former is of very much lower level than the latter, it will always be discriminated against; it is only when its level is allowed to rise near to that of the synchronizing signal that discrimination is likely to be reversed—but exact calculation of the effect is not easy. A fuller discussion of this particular point is given on page 116 of the fourth part of Reference 5.

The measured suppression characteristic for the *RC* oscillator is shown with that of the Wien-bridge oscillator in Fig. 8. The additional frequency discrimination obtained with the *RC* oscillator due to its higher effective Q-factor is clearly shown, and the divergence of the curves suggests different non-linear laws.

An estimate of the frequency response of the oscillator to an unwanted tone accompanying the synchronizing tone can be made by assuming a completely linear circuit, and, while putting in the conditions for an oscillatory loop (i.e. initial loop gain of unity) ignoring the effects of oscillation and synchronization.

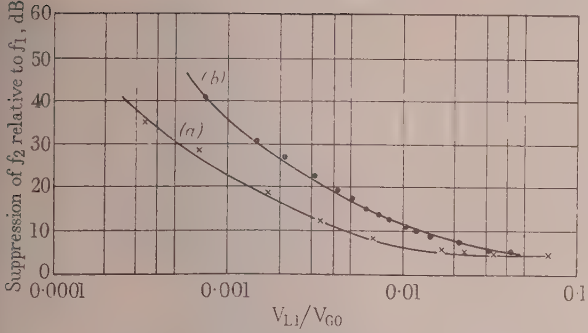


Fig. 8.—Comparison of the suppression curves for the Wien-bridge and *RC* oscillators.

$f_1 = f_0; f_1 - f_2 = 100$ c/s; $V_{L2}/V_{L1} = 0{\cdot}1$
(*a*) Wien-bridge oscillator.
(*b*) *RC* oscillator.

The use of this calculation in association with the calculations of non-linear discrimination for $f_2 \simeq f_1 = f_0$ to obtain an overall response has been discussed elsewhere.[6] The gain to the unwanted frequency is given by

$$\left|\frac{V_{G2}}{V_{L2}}\right| = \sqrt{\left[\frac{1 + Q^2(1 - x^2)^2}{(1 - x)^2 + Q^2(1 - x^2)^2}\right]} \quad . \quad . \quad . \quad (5)$$

where $x = f_2/f_1$. This gives, in effect, the gain due to regeneration (or positive feedback), and a calculated curve for the Wien-bridge oscillator is given in Fig. 9 which was derived from the above equation for $Q = 1/3$. The measured curve for the oscillator is also shown. This measurement is, of course, not a measurement of discrimination relative to the synchronizing tones but merely one of frequency response. It will be seen that the agreement of frequency response is good except at small values of $f_2 - f_0$, where the effects of non-linearity begin seriously to invalidate the calculations.

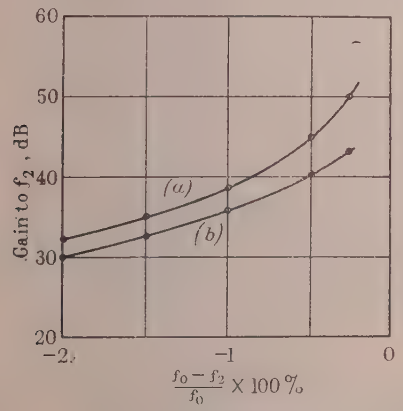


Fig. 9.—Gain to the unwanted frequency (Wien-bridge oscillator).
(*a*) Calculated. (*b*) Measured.

It is clearly difficult in practice to separate completely the effects of frequency and non-linear discrimination. The results given do show, however, the reality and considerable magnitude of the non-linear discrimination.

(2.3) Response to Synchronizing Signal mixed with Noise

The theory from which eqn. (3) was derived[6] can be extended to the case of noise accompanying a predominant synchronizing signal on the same assumptions that were made in the original analysis, namely that none of the components of the applied spectrum suffers any amplitude-frequency discrimination or any phase-shift in the oscillatory loop circuit. This restriction implies in practice that the natural frequency and the synchronizing frequency must be almost identical, and that the Q-factor of the frequency selective element in the feedback path must be very low. The mathematical work is given in Section 3, from which it will be seen that the basic relation corresponding to eqn. (3) is not readily expressible in explicit form. However, the nature of the relationship is discussed in Section 3 and shown graphically in Fig. 12, from which it can be seen that the behaviour is of the same nature as that obtained with a single interfering tone, and only slightly smaller in magnitude.

Table 2

Improvement of Signal/Noise Ratio in Synchronized *RC* Oscillator (Natural and Synchronized Frequencies Equal)

Frequency (increments in c/s)	Signal/noise ratio = 10 dB $V_{L1}/V_{G0} = 0{\cdot}0037$		Signal/noise ratio = 0 dB $V_{L1}/V_{G0} = 0{\cdot}0027$	
	Input	Output	Input	Output
$f_0 + 100$	0·010	—	0·028	0·018
$f_0 + 50$	0·017	0·005	0·048	0·024
$f_0 + 20$	0·036	0·024	0·10	0·147
f_0 (10 kc/s)	1	1	1	1
$f_0 - 20$	0·043	0·018	0·12	0·076
$f_0 - 50$	0·014	0·003	0·04	0·012
$f_0 - 100$	0·009	—	0·024	—

Table 2 gives experimental results for the suppression of a narrow band of noise (10 c/s between 3 dB points) accompanying the synchronizing signal and symmetrically disposed about it. The measurements, which were taken with the *RC* oscillator, have been normalized to unity output of the wanted frequency (in this case f_0). It should be pointed out that for practical reasons the measurements are made at such frequency intervals that the central part of the applied noise band is not examined. At the higher signal/noise ratios the suppression in this band should be no different from that in the outer parts of the spectrum, but at lower ratios (say 10 dB downwards) cross-products are formed between the noise and the synchronizing tone which cause the suppression in the outer parts to be less than that in the central part of the band. The noise suppression appears to be about 10 dB when the signal/noise ratio is +10 dB, and is about the value to be expected from Fig. 8 if allowance is made for frequency selectivity and the lower suppression predicted by the theory. Greater suppression could undoubtedly be obtained by reducing V_{L1}/V_{G0}, following the same sort of law as in Fig. 8, but it was difficult to make satisfactory measurements under these conditions. Of course, in practice, the input level must be high enough to keep the oscillator synchronized over a reasonable range of frequency difference $(f_0 - f_1)$, so a high degree of suppression of the noise requires a very stable oscillator.

Decreasing the signal/noise ratio is shown to reduce the suppression, despite the more favourable value of V_{L1}/V_{G0}. At this signal/noise ratio the assumption that $nx^2 \ll 1$ in eqn. (21) is no longer valid.

Results for suppression of the noise when there is a frequency difference of 50 c/s between the synchronizing signal and the natural oscillation are given in Table 3. For this value of

Table 3

Improvement of Signal/Noise Ratio in Synchronized *RC* Oscillator (Natural and Synchronized Frequencies Unequal)

N.B. In the absence of noise, synchronization is obtained with $V_{L1}/V_{G0} = 0{\cdot}009$.

Frequency (increments in c/s)	Signal/noise ratio = 10 dB					Signal/noise ratio = 0 dB				
	Input spectrum	V_{L1}/V_{G0}				Input spectrum	V_{L1}/V_{G0}			
		0·0037	0·0074	0·018	0·036		0·0027	0·0054	0·0135	0·027
$f_0 + 50$ $f_0 + 30$ $f_0 + 10$	0·01 0·013 0·015	0·58 1·67 0·75	0·013 0·020 0·030	— 0·003 0·003	— — —	0·028 0·035 0·044	0·22 0·37 0·50	0·10 0·13 0·25	— 0·022 0·044	— 0·02 0·03
f_0 (10 050 c/s)	0·017	2·08	0·043	0·005	0·006	0·048	1·5	0·40	0·056	0·035
$f_0 - 10$ $f_0 - 20$ $f_0 - 30$ $f_0 - 40$	0·021 0·029 0·036 0·071	12·5 0·67 0·67 0·58	0·057 0·067 0·10 0·10	0·007 0·012 0·021 0·045	0·007 0·011 0·019 0·050	0·060 0·08 0·10 0·20	2·0 0·63 0·38 0·50	0·50 0·70 0·50 0·40	0·056 0·11 0·14 0·28	0·050 0·075 0·100 0·20
$f_0 - 50 = f_1$	1	1	1	1	1	1	1	1	1	1
$f_0 - 60$ $f_0 - 70$ $f_0 - 80$ $f_0 - 100$	0·071 0·043 0·029 0·014	0·92 0·21 0·13 0·12	0·067 0·023 0·012 0·005	0·045 0·024 0·012 0·008	0·075 0·025 0·017 0·008	0·20 0·12 0·08 0·04	0·35 0·18 0·075 0·038	0·35 0·60 0·11 0·06	0·22 0·11 0·072 0·039	0·25 0·11 0·075 0·040

$f_0 - f_1$ reference to Fig. 2 shows that V_{L1}/V_{G0} must be at least 0·009 to effect synchronization. It will, in fact, have to be larger than this, as the addition of noise to the synchronizing signal may be considered as being equivalent to a complex modulation which includes modulation of the signal envelope by the noise, and from previous work[4] the frequency range over which the oscillator may be synchronized will be determined by the minimum amplitude of the synchronizing signal, i.e. the value at the lowest trough. As V_{L1}/V_{G0} nears the critical value (say 0·0074 in Table 3) the oscillator will alternate between the synchronized and unsynchronized conditions, but when not locked to the synchronizing tone its free-oscillation frequency will be modified to an extent depending on the level to which the input falls. (This is the effect shown in Fig. 3.) An output spectrum extending over the frequency difference $(f_0 - f_1)$ and having a relatively large amplitude may then be expected, and this is confirmed in Table 3. When the oscillator is synchronized ($V_{L1}/V_{G0} = 0{\cdot}018$ in Table 3) the noise is suppressed by about 6 dB, and this is about the expected value for this ratio of V_{L1}/V_{G0}.

The reduced suppression due to the decreased signal/noise ratio is again apparent.

(2.4) Response to Injected Noise without Coherent Signal

Table 4 gives measured results when noise alone was fed to the oscillator. The difference frequency $(f_0 - f_1)$ was 40 c/s, with f_1 taken at the centre of the noise band. It clearly shows that the free oscillation is not suppressed until V_{L1}/V_{G0} has reached values far too high to obtain non-linear suppression of the frequencies on either side of f_1. In all cases, the output

Table 4

Injection of Noise without Coherent Signal into *RC* Oscillator. Natural and Noise Centre Frequencies Unequal. Noise Bandwidth 10 c/s. $V_{G0} = 300$ mV

Frequency (increments in c/s)	Input spectrum	Noise input (mV)								
		0·5	1	2	3	5	10	20	50	100
$f_0 + 100$ $f_0 + 50$ $f_0 + 20$	0·015 0·035 0·062	0·13 1·17 0·83	0·11 0·53 1·0	0·062 0·25 0·50	0·031 0·1 0·25	0·028 0·078 0·17	0·030 0·070 0·15	0·029 0·058 0·11	0·025 0·057 0·093	0·033 0·067 0·093
f_0 (10 040 c/s)	0·10	16·7	4·66	1·5	0·5	0·28	0·2	0·15	0·14	0·15
$f_0 - 20$	0·22	1·67	2·67	1·12	0·62	0·44	0·3	0·25	0·21	0·3
$f_0 - 40 = f_1$	1	1	1	1	1	1	1	1	1	1
$f_0 - 50$ $f_0 - 60$ $f_0 - 80$ $f_0 - 100$	0·4 0·17 0·11 0·056	1·17 0·5 0·12 —	1·0 0·4 0·1 0·053	1·25 0·37 0·15 0·062	0·75 0·44 0·10 0·050	0·78 0·44 0·11 0·067	0·5 0·3 0·130 0·070	0·5 0·25 0·1 0·053	0·36 0·25 0·13 0·071	0·4 0·27 0·13 0·080

spectrum is wider than the input spectrum. The use of the synchronized oscillator to change the shape of a noise spectrum (i.e. to make it narrower), while theoretically feasible when the input noise spectrum is centred at its natural frequency, is clearly of little practical value.

(3) THEORY OF THE REGENERATIVE CUBE-LAW CIRCUIT (i.e. A SYNCHRONIZED OSCILLATOR) WITH THE SYNCHRONIZING TONE ACCOMPANIED BY INTERFERING TONES OR NOISE

The arrangement of a generalized regenerative cube-law circuit is shown in Fig. 10, and the basic equation for equilibrium is evidently

$$v_1 = v_L + v_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

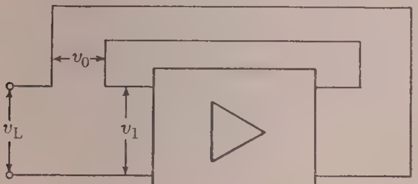


Fig. 10.—Regenerative cube-law circuit.
$v_0 = a_1v_1 + a_2v_1^2 + a_3v_1^3$

where v_L = Applied signal.
v_0 = Fed-back signal.
v_1 = Signal applied to cube-law device.

The cube law referred to is

$$v_0 = a_1v_1 + a_2v_1^2 + a_3v_1^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

The behaviour of the cube-law device without any regeneration must first be established. If

$$v_1 = V_1[\cos\omega_p t + x_1\cos(\omega_p - \omega_{q1})t + \ldots + x_n\cos(\omega_p - \omega_{qn})t] \quad . \quad . \quad . \quad . \quad (8)$$

then v_0 as given by eqns. (7) and (8) contains many kinds of output component, such as direct current, envelope components, high harmonics of ω_p and $(\omega_p - \omega_q)$, and so on; but as narrow-band systems are concerned in practice, assume that all $\omega_q \ll \omega_p$ and that the only output components of interest are those around p. They are given by

$$\begin{aligned}
v_0 = {} & \left[a_1V_1 + a_3V_1^3\left(\tfrac{3}{4} + \tfrac{3}{2}\sum_{r=1}^{n} x_r^2\right)\right]\cos\omega_p t \\
& + \sum_{r=1}^{n}\left\{a_1V_1x_r + a_3V_1^3\left[\tfrac{3}{4}x_r^3 + \tfrac{3}{2}x_r\left(1 + \sum_{\substack{m=1 \\ \text{but } m \neq r}}^{n} x_m^2\right)\right]\right\}\cos(\omega_p - \omega_{qr})t \\
& + \sum_{r=1}^{n} a_3V_1^3 \times \tfrac{3}{4}x_r\cos(\omega_p + \omega_{qr})t \\
& + \sum_{r=1}^{n} a_3V_1^3 \times \tfrac{3}{4}x_r^2\cos(\omega_p - 2\omega_{qr})t \\
& + \sum_{r=1}^{n-1}\sum_{m=r+1}^{n} a_3V_1^3 \times \tfrac{3}{2}x_rx_m[\cos(\omega_p - \omega_{qr} + \omega_{qm})t \\
& \qquad + \cos(\omega_p + \omega_{qr} - \omega_{qm})t + \cos(\omega_p - \omega_{qr} - \omega_{qm})t] \\
& + \sum_{r=1}^{n-2}\sum_{m=r+1}^{n-1}\sum_{s=m+1}^{n} a_3V_1^3 \times \tfrac{3}{2}x_rx_mx_s[\cos(\omega_p + \omega_{qr} - \omega_{qm} - \omega_{qs})t \\
& + \cos(\omega_p - \omega_{qr} + \omega_{qm} - \omega_{qs})t + \cos(\omega_p - \omega_{qr} - \omega_{qm} + \omega_{qs})t]
\end{aligned} \quad . \quad . \quad . \quad . \quad (9)$$

If $V_1\cos\omega_p t$ be regarded as a coherent signal but the remainder of the tones as a representation of noise, then all x's are made equal and n is made infinite. To evaluate eqn. (9) in these circumstances the number of components in each group of terms must be known. [Note that $(\theta_1 + \theta_2 + \theta_3 + \ldots + \theta_r + \ldots + \theta_n)^3$

$$= \underbrace{\sum_{r=1}^{n}\theta_r^3}_{n \text{ terms}} + \underbrace{\sum_{r=1}^{n}\sum_{\substack{m=1 \\ (r \neq m)}}^{n} 3\theta_r\theta_m^2}_{n(n-1) \text{ terms}} + \underbrace{\sum_{r=1}^{n-2}\sum_{m=r+1}^{n-1}\sum_{s=m+1}^{n} 6\theta_r\theta_m\theta_s}_{\frac{1}{6}n(n-1)(n-2) \text{ terms}}$$

This expansion is used as the basis for eqn. (9).] It can be seen that there are n terms in each of the groups $(\omega_p - \omega_{qr})$, $(\omega_p + \omega_{qr})$ and $(\omega_p - 2\omega_{qr})$, there are $\frac{3}{2}n(n - 1)$ components involving two different ω_q's, and there are $\frac{1}{2}n(n - 1)(n - 2)$ components involving three different ω_q's. So the output from the cube-law circuit is

$$\text{signal} = \left[a_1V_1 + a_3V_1^3\left(\frac{3}{4} + \frac{3}{2R_1^2}\right)\right]\cos\omega_p t \quad . \quad (10)$$

where R_1 is the input signal/noise (r.m.s.) ratio, and

$$\underset{\text{r.m.s.)}}{\text{noise}} = V_{n1}\left[a_1^2 + 3a_1a_3V_1^2 + \frac{45}{16}a_3^2V_1^4\left(1 + \frac{2\cdot8}{R_1^2} + \frac{1\cdot2}{R_1^4}\right)\right]^{\frac{1}{2}} . \quad (11)$$

where V_{n1} is the input r.m.s. noise voltage.

It can be seen from eqns. (10) and (11)—and putting in numerical values will quickly confirm—that if a_3/a_1 is negative, as is usual, then the output signal/noise ratio in the band around ω_p (i.e. the through-transmission band) is higher than the input ratio R_1 for small values of V_1, reaching a maximum before falling to zero, and later recovering somewhat as V_1 is raised to large values. A typical graph of the relationship is shown in Fig. 11. It would prove difficult in practice, however, to make any real use of this performance as it stands; it demands a perfect cube law for the maximum peak to be achieved, and this is at one critical signal level for a particular noise level. But the use of regeneration, as in Fig. 10, permits a high signal/noise discrimination to be achieved in a (theoretically, at least) more attractive way.

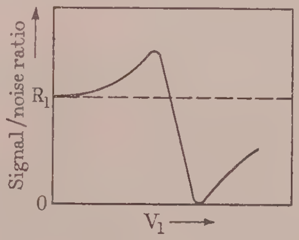


Fig. 11.—Variation of signal/noise ratio in through-transmission band of cube-law circuit.

When regeneration is applied to the cube-law circuit, the possibility of self-oscillation must be considered. It will be shown later that the most useful condition is when the regeneration is large enough to permit oscillation, and therefore for stable operation the system must take the form of a synchronized oscillator, typically as shown in Fig. 1. The feedback is tuned or filtered so that oscillation, if it occurs, can only be of frequency around ω_p, and the applied signal acts as a synchronizing signal to suppress any free oscillation. Assume that the feedback filtration has a flat amplitude-response and zero phase-shift over the band considered. Then using the symbols V_G, V_L and V_F in place of v_1, v_L and v_0 in eqn. (6), the general relation

$$V_G = V_L + V_F \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

holds. Now the relationship between V_F and V_G is the same as that between v_0 and v_1 in eqns. (8) and (9). Using this, V_L can be obtained as a function of V_G. The working below is not rigorous, but is quite adequate for ordinary purposes. Thus using suffixes to indicate the frequencies concerned, it follows from eqns. (9) and (12) that

$$(a) \qquad V_F(p) = a_1 V_G(p) + a_3[V_G(p)]^3\left(\tfrac{3}{4} + \tfrac{3}{2}\sum_{r=1}^{n} x_r^2\right) \quad .$$

Therefore

$$V_L(p) = V_G(p)\left\{1 - a_1 - a_3[V_G(p)]^2\left(\tfrac{3}{4} + \tfrac{3}{2}\sum_{r=1}^{n} x_r^2\right)\right\} \quad . \quad (13)$$

$$(b) \quad V_F(p - q_r) = a_1 x_r V_G(p) + a_3[V_G(p)]^3\left[\tfrac{3}{4}x_r^3 + \tfrac{3}{2}x_r\left(1 + \sum_{\substack{m=1 \\ \text{but } m \neq r}}^{n} x_m^2\right)\right]$$

(where the ratios x refer to the grid voltage V_G).

Therefore $V_L(p - q_r)$

$$= x_r V_G(p)\left\{1 - a_1 - a_3[V_G(p)]^2\left[\tfrac{3}{4}x_r^2 + \tfrac{3}{2}\left(1 + \sum_{\substack{m=1 \\ \text{but } m \neq r}}^{n} x_m^2\right)\right]\right\} \quad . \quad . \quad . \quad . \quad (14)$$

$$(c) \qquad V_F(p + q_r) = \tfrac{3}{4}x_r a_3[V_G(p)]^3$$

But this frequency is one not contained in the input V_L, which is assumed, for clarity only, to contain only $(\omega_p - \omega_q)$ frequencies. The frequencies $\omega_p + \omega_q$ are generated by the non-linearity of the circuit. It follows that

$$V_L(p + q_r) = 0 \text{ and } V_G(p + q_r) = V_F(p + q_r) = \tfrac{3}{4}x_r a_3[V_G(p)]^3 \quad . \quad . \quad . \quad . \quad (15)$$

(*d*) Similarly $V_L(p \pm q_r \pm q_m) = 0$ and

$$V_G(p \pm q_r \pm q_m) = V_F(p \pm q_r \pm q_m) = \tfrac{3}{2}x_r x_m a_3[V_G(p)]^3 \quad . \quad . \quad . \quad . \quad (16)$$

(*e*) Similarly $V_L(p \pm q_r \pm q_m \pm q_s) = 0$ and

$$V_G(p \pm q_r \pm q_m \pm q_s) = V_F(p \pm q_r \pm q_m \pm q_s) = \tfrac{3}{2}x_r x_m x_s a_3[V_G(p)]^3 \quad . \quad (17)$$

(*f*) The terms in $\omega_p - 2\omega_q$ can be similarly dealt with, but are omitted here as they make no finite contribution to the noise power.

There are also additional frequencies introduced, because of further non-linear action on the frequencies discussed under (*c*), (*d*) and (*e*) above, these being fed back to the grid; and so on, *ad infinitum*. Fortunately, the effect can safely be neglected, since, if $nx^2 \ll 1$ as it usually must be for the system to be useful without frequency discrimination, the contribution of these further frequencies to the overall noise power is negligible.

$$\frac{R_G}{R_L} \simeq \frac{1 - a_1 - \tfrac{3}{2}a_3[V_G(p)]^2\left(1 + \dfrac{1}{R_G^2}\right)}{\left\{1 - a_1 - \tfrac{3}{4}a_3[V_G(p)]^2\left(1 + \dfrac{2}{R_G^2}\right)\right\}\sqrt{\left\{1 + \tfrac{9}{16}a_3^2[V_G(p)]^4\left(1 + \dfrac{6}{R_G^2} + \dfrac{2}{R_G^4}\right)\right\}}} \quad . \quad (24)$$

This gives the relationship between input and grid voltage for complex tone inputs. The output may be taken from the grid, in which case these relationships apply directly, or it may be taken from the anode, in which case the effects of a further passage through the cube-law device must be considered; this latter is not usually of great significance, since the system is used, as will be seen later, at low input levels.

It can be seen that, if there is only one interfering tone accompanying the synchronizing tone, the ratio of interfering to synchronizing tone at the grid is x, and at the input it is

$$\frac{xV_G(p)\{1 - a_1 - \tfrac{3}{2}a_3[V_G(p)]^2(1 + x^2)\}}{V_G(p)\{1 - a_1 - \tfrac{3}{4}a_3]V_G(p)]^2(1 + 2x^2)\}} \quad . \quad . \quad (18)$$

Therefore the ratio by which the interfering tone is suppressed relative to the synchronizing tone is

$$\frac{1 - a_1 - \tfrac{3}{2}a_3[V_G(p)]^2}{1 - a_1 - \tfrac{3}{4}a_3[V_G(p)]^2} \quad . \quad . \quad . \quad . \quad (19)$$

where it is assumed that $x \ll 1$.

It can be seen that if $a_1 = 1$ (i.e. loop gain is unity when there is no signal), the ratio of suppression is 2, i.e. the interfering tone is reduced 6 dB relative to the synchronizing frequency. If $a_1 > 1$, which is the normal condition, and since a_3 is negative in any useful practical circuit, the suppression is greater than 2, reaching very high values when the injected voltage, V_L, is very small, since $V_G(p)$ is then very little greater than the free oscillation amplitude, V_{G0}, which is given by

$$V_{G0} = \sqrt{\frac{1 - a_1}{\tfrac{3}{4}a_3}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

and the denominator of eqn. (19) approaches zero.

When the tone ω_p is accompanied by noise, all x's are made equal and n is made infinite as before. Thus the input signal/noise ratio can be expressed as

$$R_L = \frac{V_L(p)}{\sqrt{(n)}V_L(p - q_r)}$$

$$= \frac{1 - a_1 - a_3[V_G(p)]^2(\tfrac{3}{4} + \tfrac{3}{2}nx^2)}{\sqrt{(n)}x\{1 - a_1 - a_3[V_G(p)]^2[\tfrac{3}{4}x^2 + \tfrac{3}{2}(1 + nx^2)]\}} \quad . \quad (21)$$

Since, for synchronization to be maintained, the tone ω_p should predominate, it may be assumed for purposes of approximation in subsidiary terms only that $nx^2 \simeq 1/R_G^2$, where R_G is the signal/noise ratio on the grid. Thus

$$R_L \simeq \frac{1 - a_1 - a_3[V_G(p)]^2\left(\dfrac{3}{4} + \dfrac{3}{2R_G^2}\right)}{\sqrt{(n)}x\left\{1 - a_1 - a_3[V_G(p)]^2\left(\dfrac{3}{2} + \dfrac{3}{2R_G^2}\right)\right\}} \quad . \quad (22)$$

Similarly

$$R_G \simeq 1/[\sqrt{(n)}x]\left\{1 + \tfrac{9}{16}a_3^2[V_G(p)]^4\left(1 + \frac{6}{R_G^2} + \frac{2}{R_G^4}\right)\right\}^{\frac{1}{2}} \quad .2(3)$$

The ratio of improvement of signal/noise ratio between the input and the grid is therefore

Now the form of these results is inconvenient, as it is in terms of $V_G(p)$, which is not the known input signal. It would be much better to have it in terms of $V_L(p)$, which is what is given in a practical problem. But the relationship between $V_L(p)$ and $V_G(p)$ is given by eqn. (13), where Σx^2 can now be replaced by $1/R_G^2$, and it is clear that $V_G(p)$ cannot readily be expressed as an explicit function of $V_L(p)$ but is most readily determined from a family of curves relating $V_L(p)$ and $V_G(p)$ for various values of a_1 and a_3/a_1; R_G also is involved, but if large does not seriously influence this stage of the calculation.

Although the numerical calculations are thus rather involved, it is quite easy to deduce the general behaviour of the system from eqn. (24). The improvement in signal/noise ratio depends on the actual magnitudes of a_1 and a_3 and not only on their ratio. Assume for simplicity that $a_3[V_G(p)]^2$ is small compared with unity, as it is in the most useful practical cases, and that R_G is large. Then consider three ranges of a_1 (the loop gain without any signal) as follows:

(i). $a_1 < 1$. Here R_G/R_L is never large, and when $a_1 \ll 1$ or $V_G(p)$ is very small then R_G/R_L approximates to unity. Elsewhere, if a_3 is negative, R_G/R_L is slightly in excess of unity.

(ii) $a_1 \simeq 1$. This is the condition where, in the absence of signal, free oscillation could just commence. Here $R_G/R_L \simeq 2$, so that 6 dB improvement is obtained in signal/noise ratio.

(iii) $a_1 > 1$. This is the synchronized-oscillation range, and best results are obtained when $V_L(p)$ is very small. In such a case, $V_G(p)$ approximates to a value $\sqrt{[(1-a_1)/\frac{3}{4}a_3]}$ and the denominator of eqn. (24) approaches zero. The improvement in signal/noise ratio is then very great.

It will be seen that the general behaviour is the same for noise as for single interfering tones, only the numerical values being different. Fig. 12 shows the behaviour in graphical form. The curves are of much the same nature whether the abscissa is the grid voltage (V_G) or the injected voltage (V_L), but as the relationship between V_G and V_L is very non-linear, the actual shapes of the curves are different on the two scales.

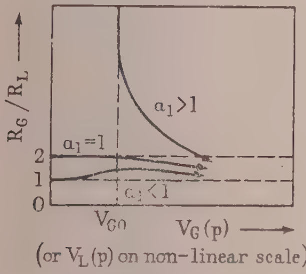


Fig. 12.—Improvement in signal/noise ratio in regenerative cube-law device.

(3.1) Theory of the Synchronized Oscillator with the Synchronizing Tone Amplitude- or Phase-Modulated by an Interfering Signal

The work in Section 3 was quite general, but when the synchronizing tone is modulated by the interference, rather special results are obtained which, although deducible from Section 3, are more conveniently and more clearly derivable by a direct approach. Modulation of the cos $\omega_p t$ tone by an interference cos $\omega_q t$ means that a pair or system of side tones that have a special relationship is applied to the non-linear circuit; but the work of Section 3 assumed that all tones were unrelated.

It will be assumed that the interfering modulation is sinusoidal; for complex modulating signals the work can readily be extended as in Section 3.

(3.1.1) Amplitude-Modulated Synchronizing Tone.

Considering first the cube-law circuit without regeneration, the input signal can be written

$$v_1 = V_1(1 + m\cos\omega_q t)\cos\omega_p t \quad . \quad . \quad . \quad (25)$$

Then the output in the through-transmission band, i.e. centred around the input band ($\omega_p \pm \omega_q$), is easily shown to be

$$\begin{aligned} v_0 = V_1\Big\{&a_1 + \tfrac{3}{4}a_3V_1^2(1 + \tfrac{3}{2}m^2) \\ &+[a_1 + \tfrac{9}{4}a_3V_1^2(1 + \tfrac{1}{4}m^2)]m\cos\omega_q t \\ &+ \tfrac{9}{8}a_3V_1^2m^2\cos 2\omega_q t + \tfrac{3}{16}a_3V_1^2m^3\cos 3\omega_q t\Big\}\cos\omega_p t \quad . \ (26) \end{aligned}$$

Since a_3 is usually negative, it is seen that the modulation in the output is less than in the input, especially if m is small.

For the circuit with regeneration the same non-rigorous arguments are adopted as in Section 3 by assuming that a signal $V_G(1 + m\cos\omega_q t)\cos\omega_p t$ exists at the grid, although accompanied by other frequencies produced by non-linearity and returned to the grid by the feedback path. Provided $m^2 \ll 1$, the effect of these other frequencies on the fundamental modulation component is negligible.

Then the fed-back voltage, V_F, is given by eqn. (7) with V_F replacing v_0 and V_G replacing V_1. Therefore, since $V_L = V_G - V_F$, it follows that

$$\begin{aligned} V_L = V_G\Big\{&1 - a_1 - \tfrac{3}{4}a_3V_G^2(1 + \tfrac{3}{2}m^2) \\ &+ [1 - a_1 - \tfrac{9}{4}a_3V_G^2(1 + \tfrac{1}{4}m^2)]m\cos\omega_q t\Big\}\cos\omega_p t \quad . \quad (27) \end{aligned}$$

ignoring the effect of second and third-harmonic modulation.

Thus, considering only the fundamental component of the modulation, the effective depth of modulation at the input is evidently

$$m_L = \frac{1 - a_1 - \frac{9}{4}a_3V_G^2(1 + \frac{1}{4}m^2)}{1 - a_1 - \frac{3}{4}a_3V_G^2(1 + \frac{3}{2}m^2)}m \quad . \quad . \quad . \quad (28)$$

so that the ratio of reduction in the depth of modulation between input and grid is

$$\frac{m_L}{m} \simeq \frac{1 - a_1 - \frac{9}{4}a_3V_G^2}{1 - a_1 - \frac{3}{4}a_3V_G^2} \quad . \quad . \quad . \quad . \quad (29)$$

since it is assumed that $m^2 \ll 1$. This can be compared with eqn. (19) for a single interfering tone accompanying the synchronizing tone, and it will be seen that the difference is that the suppression is greater for modulation. For example, when $a_1 = 1$ (i.e. incipient oscillation in the absence of signal), $m_L/m = 3$, whereas for a single tone the suppression was only 2. This is easily explained in terms of Section 1, since the side tone ($\omega_p - \omega_q$) is reduced (assuming, as usual, that a_3 is negative) by the non-linear image tone [eqn. (15)] of the ($\omega_p + \omega_q$) side tone, and vice versa.

(3.1.2) Phase-Modulated Synchronizing Tone.

If the input signal to the cube-law circuit is phase-modulated by an interfering frequency, it can be written

$$v_1 = V_1\cos(\omega_p t + m\sin\omega_q t) \quad . \quad . \quad . \quad (30)$$

When this is substituted in the non-linear law, eqn. (7), it can be seen that the phase-modulation does not affect the behaviour, and the output in the through-transmission band is

$$v_0 = (a_1V_1 + \tfrac{3}{4}a_3V_1^3)\cos(\omega_p t + m\sin\omega_q t) \quad . \quad . \quad (31)$$

since

$$\begin{aligned} \cos^3(\omega_p t + m\sin\omega_q t) &=. \tfrac{3}{4}\cos(\omega_p t + m\sin\omega_q t) \\ &+ \tfrac{1}{4}\cos 3(\omega_p t + m\sin\omega_q t) \quad . \quad (32) \end{aligned}$$

Thus when the regenerative circuit is considered it is seen that the modulation index is unaffected by the circuit, and no suppression of the modulation is obtained. This is confirmed by the fact that synchronized oscillators have been used as amplitude-limiters for phase- and frequency-modulation receivers.[7] The behaviour to phase-modulation is thus different from that to any other interfering signal.

This behaviour is, however, entirely consistent with the work in Section 3, and can be illustrated in terms of Section 3 thus. The carrier and first-order side tones of eqns. (30) are

$$V_1\{J_0(m)\cos\omega_p t - J_1(m)[\cos(\omega_p - \omega_q)t - \cos(\omega_p + \omega_q)t]\} \quad . \quad . \quad . \quad . \quad . \quad (33)$$

where $J_0(m)$ and $J_1(m)$ are Bessel functions. Assume $m^2 \ll 1$; then $J_0(m) \simeq 1$ and $J_1(m) \simeq m/2$. Considering the non-linear image frequencies, $\omega_p + \omega_q$ produced from $\omega_p - \omega_q$ and vice versa [see eqns. (9) and (15)], the resultant amplitude of each side tone in the output of the cube-law circuit is approximately

$$\frac{m}{2}V_1[a_1 + a_3V_1^2(\tfrac{3}{2} - \tfrac{3}{4})] = \frac{m}{2}V_1(a_1 + \tfrac{3}{4}a_3V_1^2) \quad . \quad (34)$$

where the $\frac{3}{2}$ is reduced to $\frac{3}{4}$ owing to the image tone of opposite polarity. Since the output of carrier is

$$V_1(a_1 + \tfrac{3}{4}a_3V_1^2)$$

it is clear that the whole basis of discrimination (i.e. the coefficient $\frac{3}{2}$ for the interference and $\frac{3}{4}$ for the synchronizing tone) has disappeared due to the special relationships of phase-modulation.

(4) CONCLUSIONS

The theoretical part of the paper has shown the origin of the non-linear discrimination of a synchronized oscillator against interfering signals accompanying the synchronizing signal, and the differences in performance when the interference modulates the synchronizing signal have been made clear. The experimental work described shows clearly the physical reality of the effect, and to the extent permitted by difficulties of measurement, particularly of system parameters (such as the coefficients a_1 and a_3), there is good agreement between theory and experiment. The conclusion reached is that the use of a synchronized oscillator as a highly-selective circuit with a centre frequency automatically adjusting itself to the required band is quite feasible so long as a coherent and dominant synchronizing tone is applied, i.e. that unwanted signals have a level below that of the synchronizing tone. For good selectivity utilizing the non-linear discrimination as well as the frequency discrimination, a very stable oscillator should be used with a very small voltage of the synchronizing signal, and the natural frequency should be kept as near the synchronizing frequency as possible.

If there is no dominant coherent tone applied, i.e. noise alone is injected, there is no practical benefit obtained by using a synchronized oscillator instead of a linear filter.

(5) ACKNOWLEDGMENTS

The paper is published by permission of the Admiralty, but any opinions expressed are personal to the authors. Mr. J. W. R. Griffiths gave helpful advice in the experimental part of the investigation.

(6) REFERENCES

(1) Appleton, E. V.: "Automatic Synchronization of Triode Oscillators," *Proceedings of the Cambridge Philosophical Society*, 1922–3, **21**, p. 231.

(2) Gillies, A. W.: "Electrical Oscillations," *Wireless Engineer*, 1953, **30**, p. 143.

(3) Tucker, D. G.: "Forced Oscillations in Oscillator Circuits and the Synchronization of Oscillators," *Journal I.E.E.*, 1945, **92**, Part III, p. 226.

(4) Byard, S., and Eccles, W. H.: "The Locked-in Oscillator," *Wireless Engineer*, 1941, **18**, p. 2.

(5) Tucker, D. G.: "The Synchronization of Oscillators," *Electronic Engineering*, 1943, **15**, pp. 412 and 457; 1943, **16**, pp. 26 and 114.

(6) Tucker, D. G.: "Non-linear Regenerative Circuits," *Wireless Engineer*, 1947, **24**, p. 178.

(7) Carnahan, C. W., and Kalmus, H. P.: "Synchronized Oscillators as F.M. Receiver Limiters," *Electronics*, August, 1944, p. 108.

621.372.8:621.396.11.029.6

Monograph No. 147 R
Aug. 1955

MICROWAVE PROPAGATION IN ANISOTROPIC WAVEGUIDES

By A. E. KARBOWIAK, Ph.D.

(*The paper was first received* 15th March, *and in revised form* 15th May, 1955. *It was published as an* INSTITUTION MONOGRAPH *in August*, 1955.)

SUMMARY

A detailed analysis of the propagation of electromagnetic waves in circular waveguides whose walls are anisotropic is carried out. The physical property of the wall is described in terms of "anisotropic surface impedance" and this has two principal components, which, in general, do not run along the co-ordinates of the cylinder.

It is shown that, whatever the orientation of the principal axes of the surfaces, all E- and H$_0$-modes are stable. Formulae are derived for the propagation coefficients and attenuation coefficient of the waves, and these are expressed in terms of the impedance components of the surface.

Furthermore, it is shown that all higher-order H-modes are unstable in such waveguides unless the principal axes of the surface coincide with the co-ordinate axes of the surface: a wave that is stable in an anisotropic waveguide, whatever the orientation of the axes, is a "spinning H-wave."

LIST OF PRINCIPAL SYMBOLS

r, ϕ, z = Co-ordinates of the circular cylinder.

η, ζ = Helical co-ordinates.

ψ = Lay angle of the ζ-helix (see Fig. 1).

t = Tan ψ.

μ_0, ϵ_0 = Permeability and permittivity of free space.

$k = \dfrac{2\pi}{\lambda_0} = \omega\sqrt{(\mu_0\epsilon_0)}$ = Free-space propagation coefficient.

λ_0 = Free-space wavelength.

$Z_0 = \sqrt{\left(\dfrac{\mu_0}{\epsilon_0}\right)}$ = Free-space impedance.

Z_s = Normalized (with respect to Z_0) surface impedance.

Z_η, Z_ζ = Principal "components" of Z_s.

$\gamma = \alpha + j\beta$ = Axial propagation coefficient.

α = Attenuation coefficient.

$\beta = \dfrac{2\pi}{\lambda_g}$ = Phase-change coefficient.

$h_0 = \dfrac{2\pi}{\lambda_c}$ = Cut-off coefficient of a perfect waveguide.

$h = h_0 + \delta h$ = Cut-off coefficient of an imperfect waveguide.

s = Radius of the waveguide.

m, n = Mode indices.

$\Delta\gamma$ = Wave spin coefficient.

C_v, S_v, C'_v, S'_v = Constants.

$J_m = \dfrac{J_m(hs)}{J'_m(hs)}$, where $h_0 s$ is a root of $J_m(hs) = 0$.

$J'_m = \dfrac{J'_m(hs)}{J_m(hs)}$, where $h_0 s$ is a root of $J'_m(hs) = 0$.

$p = \dfrac{\beta m}{h^2 s}$

$q = \dfrac{k}{h}$

$Z_e = \dfrac{J_m}{jq}$ = Equivalent uniform surface impedance (for E-waves).

$Z_H = jqJ'_m$ = Equivalent uniform surface impedance (for H-wave).

The rationalized M.K.S. system of units is used throughout and the time dependence exp $(j\omega t)$ is implied. Furthermore, unless otherwise stated, all field quantities are understood to contain the factor exp $(-\gamma z)$.

The wave numbers are connected by

$$\begin{aligned} k^2 &= h_0^2 + \beta^2 = h^2 - \gamma^2 \\ h &= h_0 + \delta h \\ \gamma &= j\beta + \delta\gamma = j\beta + \alpha + j\delta(\beta) \end{aligned}$$

For spinning waves we have, in addition:

$$\begin{aligned} \gamma_1 &= j\beta + \delta(\gamma_1) = j\beta + \delta\gamma' + \Delta\gamma = \gamma' + \Delta\gamma \\ \gamma_2 &= \gamma' - \Delta\gamma \end{aligned}$$

and

$$\delta\gamma' = \frac{1}{s}\frac{h^2}{k\beta}\frac{Z'/(1+t^2)}{1-\left(\dfrac{m}{hs}\right)^2}$$

$$\Delta\gamma = \frac{2pt}{s}\frac{h^2}{k\beta}\frac{(Z_\eta - Z_\zeta)/(1+t^2)}{1-\left(\dfrac{m}{hs}\right)^2}$$

The quantity Z_0 is absorbed in the symbol H (magnetic-field vector) and consequently all impedances and coupling coefficients are normalized with respect to that quantity.

(1) INTRODUCTION

Wave propagation in waveguides whose walls exhibit small but finite surface impedance has been investigated by the author elsewhere,[1] and inasmuch as some aspects of anisotropic waveguides have been touched upon, this study has for its object wave propagation in anisotropic waveguides whose principal axes of the surface do not, in general, coincide with the axes of the co-ordinate system employed. The only allied problem that has been dealt with in the past is that of propagation along helically conducting structures,[2] and this is a particular case of the general method of approach developed here.

Waveguides of anisotropic surface impedance are of moderately frequent occurrence in practical applications, and a corrugated surface or a helix[3] are just two examples of anisotropic impedance sheet. It is thought desirable to have a general method of attack on problems connected with waveguides of that nature, and the surface-impedance approach developed in Reference 1 is admirably suited for this purpose.

Accordingly, the problem of wave propagation is solved for an arbitrary surface impedance (anisotropic) and the formulae obtained relate the propagation coefficients of the wave to the dimensions of the waveguide and its surface-impedance

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Karbowiak is with Standard Telecommunication Laboratories, Ltd.

components. In the case of stable waves an expression for the equivalent homogeneous and isotropic surface impedance, Z_e, may be derived. Here, by Z_e we mean that value of uniform surface impedance for which the waveguide would have the same propagation coefficient as the investigated anisotropic waveguide.

(2) THE FORMULATION OF THE PROBLEM

Consider a cylindrical guide as shown in Fig. 1. The co-ordinates r, ϕ and z are respectively radial, circumferential and axial co-ordinates of the cylindrical guide, and the co-ordinate $r = s$ defines the guide surface, which is characterized by a uniform surface impedance Z_s.

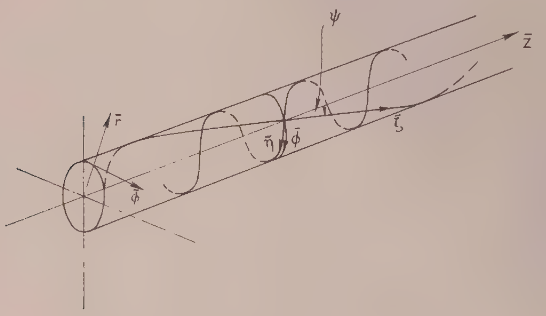


Fig. 1.—Cylindrical anisotropic waveguide.

In our present study this impedance is anisotropic and accordingly has two components Z_η and Z_ζ; these are defined by

$$Z_\eta = \frac{E_\eta}{H_\zeta}\bigg|_{r=s} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and
$$Z_\zeta = -\frac{E_\zeta}{H_\eta}\bigg|_{r=s} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The axes of η and ζ are the principal axes of the surface and in the circular guide they form two mutually orthogonal helices. The angle ψ is the lay angle of the ζ-helix.

It is convenient to develop the guide surface (as shown in Fig. 2), which can then be looked upon as a plane impedance sheet with principal rectilinear axes η and ζ. Figs. 3 and 4 are two examples of anisotropic surfaces which behave as if they were homogeneous provided that the period of heterogeneity is much smaller than the wavelength of the guided wave—an assumption it is convenient to make here, but which is by no means necessary.

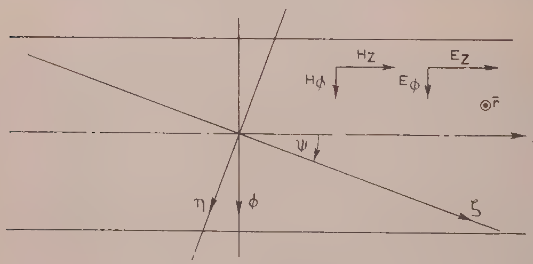


Fig. 2.—Top view of the developed waveguide surface.

Since the modes in circular guides are described in terms of the cylindrical co-ordinates r, ϕ and z, the field components E_ϕ, H_ϕ, E_z and H_z in the problems presented here must be projected along the η and ζ directions and the boundary conditions of eqns. (1) and (2) then applied. Inspection of Fig. 2 leads to the following relations between the relevant field components:

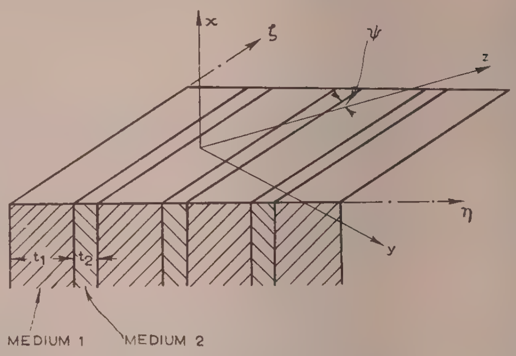


Fig. 3.—Anisotropic surface.

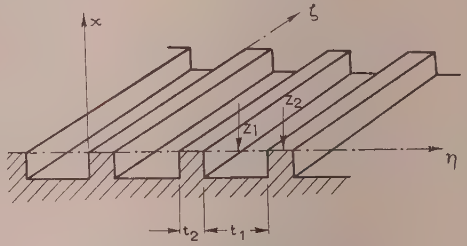


Fig. 4.—Corrugated surface.

$$\left.\begin{aligned} \begin{bmatrix} E_\eta \\ H_\eta \end{bmatrix} &= \begin{bmatrix} E_\phi \\ H_\phi \end{bmatrix} \cos\psi - \begin{bmatrix} E_z \\ H_z \end{bmatrix} \sin\psi \\ \begin{bmatrix} E_\zeta \\ H_\zeta \end{bmatrix} &= \begin{bmatrix} E_z \\ H_z \end{bmatrix} \cos\psi + \begin{bmatrix} E_\phi \\ H_\phi \end{bmatrix} \sin\psi \end{aligned}\right\} \quad . \quad . \quad (3)$$

The boundary conditions to be satisfied at the surface of the guide are obtained by substituting eqn. (3) in eqns. (1) and (2). Thus

$$-Z_\zeta = \frac{E_z + E_\phi \tan\psi}{H_\phi - H_z \tan\psi}\bigg|_{r=s} \quad . \quad . \quad . \quad . \quad (4)$$

and
$$Z_\eta = \frac{E_\phi - E_z \tan\psi}{H_z + H_\phi \tan\psi}\bigg|_{r=s} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

(3) THE SOLUTION OF THE PROBLEM

(3.1) E-Waves

An E_m-wave is derived from the wave function

$$E_z = J_m(h_0 r)\binom{\sin}{\cos} m\phi\Big) \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where $h_0 s$ are roots of $J_m(h_0 s) = 0$ and

$$k^2 = h_0^2 + \beta^2 = h^2 - \gamma^2 \quad . \quad . \quad . \quad . \quad (7)$$

In the present instance, however, owing to the finite value of the wall impedance, Z_s, the coefficient h_0 becomes $h_0 + \delta h = h$ and the field will be, in general, contaminated by other modes. Thus the total field, which is a mixture of a denumerable infinity of E- and H-modes, has the following components* on the surface of the guide:

(i) Due to E-waves

$$\left.\begin{aligned} E_z &= \sum_v J_v(hr)[C_v \cos v\phi + S_v \sin v\phi] \\ E_\phi &= \frac{j\beta}{h^2r}\sum_v vJ_v(hr)[C_v \sin v\phi - S_v \cos v\phi] \\ H_\phi &= -j\frac{k}{h}\sum_v J_v'(hr)[C_v \cos v\phi + S_v \sin v\phi] \end{aligned}\right\} \quad . \quad (8)$$

(ii) Due to H-waves

$$\left.\begin{aligned} H_z &= \sum_v J_v(hr)[C_v' \cos v\phi + S_v' \sin v\phi] \\ E_\phi &= j\frac{k}{h}\sum_v J_v'(hr)[C_v' \cos v\phi + S_v' \sin v\phi] \\ H_\phi &= j\frac{\beta}{h^2r}\sum_v vJ_v(hr)[C_v' \sin v\phi - S_v' \cos v\phi] \end{aligned}\right\} \quad . \quad (9)$$

In these equations C_v, S_v, C_v' and S_v' are constants (Fourier coefficients) and m denotes the circumferential order of the mode. Let us investigate the mode of order m. If we let $C_m = 1$ the remaining coefficients become the coupling coefficients[1] of the various modes that come into existence as a result of finite value of the wall impedance, Z_s. If, when $C_m = 1$, all the remaining coupling coefficients are small, the mode considered will be regarded as stable in the guide investigated and we can then say that the guide is supporting a substantially pure mode.[1]

To solve for h or δh and the various coupling coefficients, substitute eqns. (8) and (9) in eqns. (4) and (5), whereupon, from cross-multiplication, we get the following two simultaneous equations:

$$\begin{aligned} &\Big\{\tan\psi\,\Sigma J_v(hs)[C_v' \cos v\phi + S_v' \sin v\phi] \\ &\quad + \Sigma J_v'(hs)[C_v \cos v\phi + S_v \sin v\phi] \\ &\quad - j\frac{\beta}{h^2s}\Sigma vJ_v(hs)[C_v' \sin v\phi - S_v' \cos v\phi]\Big\}Z_\zeta \\ &\quad = \Sigma J_v(hs)[C_v \cos v\phi + S_v \sin v\phi] \\ &\quad + \tan\psi\Big\{j\frac{\beta}{h^2s}\Sigma vJ_v(hs)[C_v \sin v\phi - S_v \cos v\phi] \\ &\quad + j\frac{k}{h}Z_0\Sigma J_v'(hs)[C_v' \cos v\phi + S_v \sin v\phi]\Big\} \quad . \quad . \quad (10) \end{aligned}$$

$$\begin{aligned} \text{and } &\Big\{\Sigma J_v(hs)[C_v' \cos v\phi + S_v' \sin v\phi] \\ &\quad + \frac{j\beta}{h^2s}\Sigma vJ_v(hs)[C_v' \sin v\phi + S_v' \cos v\phi]\tan\psi \\ &\quad - j\frac{k}{h}\Sigma J_v'(hs)[C_v \cos v\phi + S_v \sin v\phi]\Big\}Z_\eta \\ &\quad = -\frac{j\beta}{h^2s}\Sigma vJ_v(hs)[C_v \sin v\phi - S_v \cos v\phi] \\ &\quad - j\frac{k}{h}Z_0\Sigma J_v'(hs)[C_v' \cos v\phi + S_v' \sin v\phi] \\ &\quad - \tan\psi\,\Sigma J_v(hs)[C_v \cos v\phi + S_v \sin v\phi] \quad . \quad . \quad (11) \end{aligned}$$

Eqns. (10) and (11) are solved for the Fourier coefficients in the usual manner. Thus, by multiplying throughout by cos $m\phi$ and then by sin $m\phi$ and integrating over the period, eqns. (10) and (11) will be found to lead to a set of four equations; these, after rearrangement of the terms and making the following simplifying substitutions

$$\left.\begin{aligned} &\frac{\beta m}{h^2s} = p; \quad \frac{k}{h} = q \\ &\frac{J_m(hs)}{J_m'(hs)} = J_m \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (12)$$

lead to

$$\left.\begin{aligned} &C_m'[J_mZ_\eta - jq] + S_m'[-jpJ_mZ_\eta \tan\psi] \\ &\qquad - C_m[J_m - jqZ_\eta]\tan\psi - S_m[jpJ_m] \\ &C_m'[jpJ_mZ_\eta \tan\psi] + S_m'[J_mZ_\eta - jq] \\ &\qquad = C_m[jpJ_m] - S_m[J_m - jqZ_\eta]\tan\psi \\ &C_m'[J_mZ_\zeta - jq]\tan\psi + S_m'[jpJ_mZ_\zeta] \\ &\qquad = C_m[J_m - jqZ_\zeta] - S_m[jpJ_m]\tan\psi \\ &-C_m'[jpJ_mZ_\zeta] + S_m'[J_mZ_\zeta - jq]\tan\psi \\ &\qquad = C_m[jpJ_m]\tan\psi + S_m[J_m - jqZ_\zeta] \end{aligned}\right\} \quad . \quad (13)$$

Since we have assumed that Z_ζ and Z_η are small, it follows that in eqns. (13)

$$\text{and} \quad \left.\begin{aligned} J_mZ_\eta &\ll jq \\ J_mZ_\zeta &\ll jq \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

and the terms can be accordingly omitted with a negligible error, particularly since J_m is small.*

The set of four equations (13) is homogeneous in the four unknowns (C_m, S_m, C_m', S_m'), and for a non-trivial solution we must have the determinant of the coefficients equal identically to zero [cf. eqn. (23) below]. This, after some lengthy but straightforward eliminations, leads to the solution:

$$J_m(1 + \tan^2\psi) = jq(Z_\zeta + Z_\eta \tan^2\psi) \quad . \quad . \quad (15)$$

Consequently the equivalent isotropic surface impedance is

$$Z_e = \frac{J_m}{jq} = \frac{Z_\zeta + Z_\eta \tan^2\psi}{1 + \tan^2\psi} \quad . \quad . \quad . \quad . \quad (16)$$

$$\text{and}^1 \qquad \delta(h) = \frac{1}{s}\frac{k}{h_0}(jZ_e) = \frac{k}{h_0}\frac{(jZ_\zeta) + (jZ_\eta)\tan^2\psi}{1 + \tan^2\psi} \quad . \quad (17)$$

In particular,[1] since $\delta(\gamma) = \alpha + j\delta(\beta)$,

$$\left.\begin{aligned} \alpha &= \frac{1}{s}\frac{k}{\beta_0}\frac{R_\zeta + R_\eta \tan^2\psi}{1 + \tan^2\psi} \\ \text{and} \qquad \delta(\beta) &= -\frac{1}{s}\frac{k}{\beta_0}\frac{X_\zeta + X_\eta \tan^2\psi}{1 + \tan^2\psi} \end{aligned}\right\} \quad . \quad . \quad . \quad (18)$$

For $\psi = 0$, $Z_e = Z_\zeta$ and for $\psi = 90°$, $Z_e = Z_\eta$, which is in agreement with results derived elsewhere.[1] If ψ is small

$$\left.\begin{aligned} &Z_e \simeq Z_\zeta(1 - \tan^2\psi) + Z_\eta \tan^2\psi \\ \text{or} \qquad &Z_e \simeq Z_\zeta + (Z_\eta - Z_\zeta)\tan^2\psi \end{aligned}\right\} \quad . \quad . \quad (18a)$$

Thus, for small values of ψ the performance of the guide depends primarily on the value of the axial component of the surface impedance: this, however, is increased by an amount proportional to the square of the lay angle ψ, and the difference between the circumferential and axial components of the surface impedance.

Since the problem has a unique solution in the form of

* Note that these are normalized with respect to Z_0; cf. list of symbols.

* $J_m = J_m(hs)/J_m'(hs)$ and (hs) is close to the root of $J_m(hs) = 0$.

eqn. (15), the stability of the wave is unimpaired by the anisotropic nature of the guide.*

(3.2) H-waves

An H_m-wave is derived from the wave function

$$H_z = J_m(h_0 r)\binom{\sin}{\cos} m\phi \quad . \quad . \quad . \quad . \quad (19)$$

where $(h_0 s)$ are roots of $J'_m(h_0 s) = 0$ and

$$k^2 = h_0^2 + \beta^2 = h^2 - \gamma^2 \quad . \quad . \quad . \quad (20)$$

But an imperfect guide cannot, in general, support a pure H_m-wave; consequently, as in connection with E-waves, we seek a solution in the form of eqns. (8) and (9). Here again, if we let $C'_m = 1$, the remaining coefficients (of the set C_v, S_v, C'_v and S'_v) become the coupling coefficients of the various modes, and these modes come into existence by virtue of the finite value of Z_s.

Since the analysis for an H-wave is very much similar to that for an E-wave, the details of the analysis will be omitted. Eqns. (10) and (11) are solved for the Fourier coefficients in the manner indicated in connection with E-waves, and this leads—for example in the case of the mth coefficients—to four simultaneous equations homogeneous in the four coefficients C'_m, S'_m, C_m and S_m. It will be convenient to make the following simplifying substitutions

$$\frac{\beta m}{h^2 s} = p; \quad \frac{k}{h} = q; \quad t = \tan\psi$$

$$J'_m = \frac{J'_m(hs)}{J_m(hs)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

and
$$Z_H = jqJ'_m = Z_e(1 + p^2) \quad . \quad . \quad . \quad . \quad (22)$$

whereupon we arrive at the following secular determinant,†

$$\begin{vmatrix} (Z_\eta - Z_H), & -jpZ_\eta t, & t, & jp \\ jpZ_\eta t, & (Z_\eta - Z_H), & -jp, & t \\ (Z_\zeta - Z_H), & jpZ_\zeta, & -1, & jpt \\ -jpZ_\zeta, & (Z_\zeta - Z_H)t, & -jpt, & 1 \end{vmatrix} = 0 \quad . \quad (23)$$

The unknown in eqn. (23) is the coefficient $h\,(= h_0 + \delta h)$, which appears implicitly in the symbol Z_H. The secular equation (23) can be reduced, after some algebraic operations, to the following quadratic equation in Z_H:

$$\begin{aligned} Z_H^2(1 + t^2)^2 - 2Z_H(1 + t^2) \\ \times [Z_\eta(1 + p^2t^2) + Z_\zeta(p^2 + t^2)] \\ + [Z_\eta(1 + p^2t^2) + Z_\zeta(p^2 + t^2)]^2 \\ - 4p^2t^2(Z_\eta - Z_\zeta) = 0 \quad . \quad (24) \end{aligned}$$

and the solution to eqn. (24) is

$$Z_H = [Z_\eta(1 \pm pt)^2 + Z_\zeta(p \mp t)^2]/(1 + t^2) \quad . \quad . \quad (25)$$

We note that for $p = 0$, we have

$$Z_{H0} = [Z_\eta + Z_\zeta t^2]/(1 + t^2) \quad . \quad . \quad . \quad . \quad (26)$$

while for $t = 0 = \psi$, eqn. (25) becomes

$$Z_H = Z_\eta + Z_\zeta p^2 \; (\psi = 0) \quad . \quad . \quad . \quad . \quad (27)$$

and for $\psi = 90°(t = \infty)$,

i.e.
$$Z_H = Z_\eta p^2 + Z_\zeta \; (\psi = 90°) \quad . \quad . \quad . \quad . \quad (28)$$

Thus, unless $p = 0$ or $\psi = 0$ or $\psi = 90°$, Z_H has two distinct values, giving separate modes. Consequently we draw the conclusion that all H-waves are unstable in anisotropic circular guides unless the principal axes of the surface coincide with the z and ϕ co-ordinate of the guide ($\tan\psi = 0$) or the wave is an H_0-wave ($p = 0$).*

If we now assume that the impedances Z_η and Z_ζ are small, by substituing eqn. (25) in eqn. (22) and expanding the Bessel functions occurring in eqn. (22) in Taylor's series about the point $(h_0 s)$, we get the following expression[1] for $\delta(h)$ $(= h - h_0)$

$$\delta(h) = \frac{1}{s}\frac{h_0}{k}\frac{jZ_H}{1 - \left(\frac{m}{h_0 s}\right)^2} \quad . \quad . \quad . \quad . \quad (29)$$

thus
$$\delta(\gamma) = \frac{1}{s}\frac{h_0^2}{k\beta}\frac{Z_H}{1 - \left(\frac{m}{h_0 s}\right)^2} \quad . \quad . \quad . \quad . \quad (30)$$

Since there are two values (Z_1 and Z_2) of Z_H given by

$$\left.\begin{aligned} Z_1(1 + t^2) = Z_\eta(1 + p^2t^2) + Z_\zeta(p^2 + t^2) + 2pt(Z_\eta - Z_\zeta) \\ Z_2(1 + t^2) = Z_\eta(1 + p^2t^2) + Z_\zeta(p^2 + t^2) - 2pt(Z_\eta - Z_\zeta) \end{aligned}\right\} (31)$$

or
$$\left.\begin{aligned} Z_1(1 + t^2) = Z' + 2pt(Z_\eta - Z_\zeta) \\ Z_2(1 + t^2) = Z' - 2pt(Z_\eta - Z_\zeta) \end{aligned}\right\} \quad . \quad . \quad (32)$$

where Z' is given by

$$Z' = R' + jX' = Z_\eta(1 + p^2t^2) + Z_\zeta(p^2 + t^2)$$

there are, therefore, two corresponding values of $\delta\gamma$ given by

$$\left.\begin{aligned} \delta(\gamma)_1 &= \frac{1}{s}\frac{h^2}{k\beta}\frac{Z'/(1 + t^2)}{1 - \left(\frac{m}{hs}\right)^2}\left(1 + 2pt\frac{Z_\eta - Z_\zeta}{Z'}\right) \\ &= \delta(\gamma)' + \Delta\gamma \\ \text{and} \quad \delta(\gamma)_2 &= \frac{1}{s}\frac{h^2}{k\beta}\frac{Z'/(1 + t^2)}{1 - \left(\frac{m}{hs}\right)^2}\left(1 - 2pt\frac{Z_\eta - Z_\zeta}{Z'}\right) \\ &= \delta(\gamma)' - \Delta\gamma \end{aligned}\right\} \quad . \quad (33)$$

where $\delta(\gamma)'$ and $\Delta\gamma$ are given by

$$\delta(\gamma)' = \frac{1}{s}\frac{h^2}{k\beta}\frac{Z'/(1 + t^2)}{1 - \left(\frac{m}{hs}\right)^2}$$

$$\Delta\gamma = \frac{2pt}{s}\frac{h^2}{k\beta}\frac{(Z_\eta - Z_\zeta)/(1 + t^2)}{1 - \left(\frac{m}{hs}\right)^2}$$

Furthermore, substitution of eqn. (31) back into the original equation yields

$$\left.\begin{aligned} \left(\frac{C'_m}{S'_m}\right)_1 = -j \\ \text{and} \quad \left(\frac{C'_m}{S'_m}\right)_2 = +j \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (34)$$

The remaining coefficients (of the set C'_v, S'_v, C_v, S_v) are, in magnitude, of the order of Z_η or Z_ζ, or less, and consequently can be neglected in comparison with C'_m and S'_m. Thus, the field in the guide may be derived from†

$$H_z = J_m(hr)[\exp(-jm\phi - \Delta\gamma z) + \exp(jm\phi + \Delta\gamma z)]\varepsilon^{-\gamma' z} \; . \; (35)$$

where γ' and $\Delta\gamma$ have meaning as defined by eqn. (33).

* See Section 8.1.

† Contrast this with the determinant of the coefficients in eqn. (13), where, provided that Z_η and Z_ζ are small, a unique solution for surface impedance is obtained.

* See Section 8.1.

† See Section 8.2.

To understand the nature of the wave as given by eqn. (35), suppose for the moment that the component impedances are pure reactances. It is then evident from eqn. (33) that $\delta(\gamma)'$ and $\Delta\gamma$ are pure imaginary quantities: if, say, $\gamma' = j\beta'$ and $\Delta\gamma = j\Delta\beta$, eqn. (35) can (neglecting a constant factor) be put into the form

$$H_z = J_m(hr)\cos(m\phi + \Delta\beta z)e^{-j\beta' z} \quad . \quad . \quad . \quad (36)$$

The wave as given by eqn. (36) is evidently an unattenuated wave whose plane of polarization is revolving clockwise while the wave is proceeding down the guide, when $\Delta\beta$ is positive; and the spin of the wave is anti-clockwise when $\Delta\beta$ is negative. Now, from eqn. (33) we observe that

$$\Delta\beta \propto 2pt(X_\eta - X_\zeta) \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

and consequently the sense of the spin of the wave is the same as the sense of the principal helix of the surface in the direction of the lower impedance. For example, if $Z_\eta > Z_\zeta$ the wave spin is in the same sense as that of the ζ-helix, but the rate of the spin of the wave is, as will be shown, much less than that of the ζ-helix.

If the guide is not loss-free the field is derived not from eqn. (36) but from

$$H_z = J_m(hr)\cosh(jm\phi + \Delta\gamma z)\varepsilon^{-\gamma' z} \quad . \quad . \quad (38)$$

and in this case the locus of the wave polarization vector, in the transverse plane of the guide, is no longer a circle but a spiral, as shown in Fig. 5. In addition, the wave suffers attenuation, and this can be calculated from eqns. (33) and (31) since

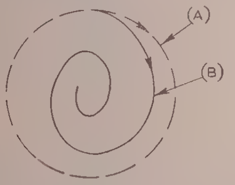


Fig. 5.—Locus of the polarization vector of the spinning H-wave.
(A) For loss-free guide.
(B) For a "lossy" guide.

$$\gamma' = \alpha + j\beta' = j\beta + \alpha + j\delta\beta' \quad . \quad . \quad . \quad (39)$$

Therefore
$$\alpha = \frac{1}{s}\frac{h^2}{k\beta}\frac{1}{1-\left(\frac{m}{hs}\right)^2}\frac{R'}{1+t^2} \quad . \quad . \quad . \quad . \quad (40)$$

and $\delta(\beta)$ is given by eqn. (40) with X' in place of R'. The quantity $\Delta\gamma$ is given by eqn. (40) with $2pt(Z_\eta - Z_\zeta)$ in place of R',

i.e.
$$\Delta\gamma = \frac{1}{sk}\frac{m/s}{1-\left(\frac{m}{hs}\right)^2}\frac{2t(Z_\eta - Z_\zeta)}{1+t^2} \quad . \quad . \quad . \quad (41)$$

In the case of the H$_0$-wave, $p = 0$ and consequently $\delta(\gamma)_1 = \delta(\gamma)_2$, while $\Delta\gamma = 0$ and therefore the wave is stable.

(4) NUMERICAL EXAMPLE

Consider a circular anisotropic guide of radius (s) 2cm, operated at 24000 Mc/s ($\lambda_0 = 1{\cdot}25$cm), and suppose that the guide surface impedance (normalized with respect to $Z_0 = 377$ ohms) has two principal components: $Z_\eta = 100Z_\zeta$ and $Z_\zeta = j \times 10^{-4}$.

In the case of any E-wave the guide will behave as if it had a uniform surface impedance given by eqn. (16); this is shown plotted as a function of the lay angle ψ in Fig. 6.

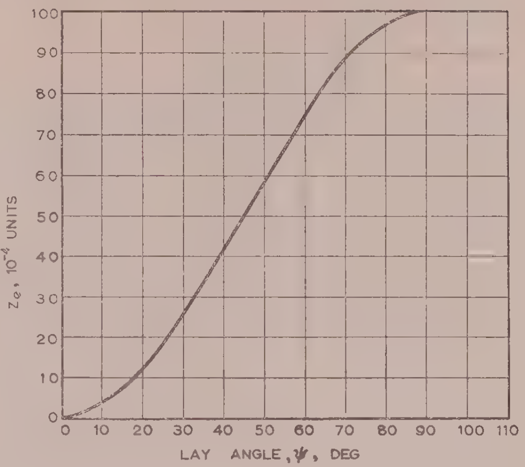


Fig. 6.—Normalized equivalent uniform surface impedance, Z_e, as a function of the lay angle, ψ, for a guide carrying an E-wave.
Components of surface impedance are: $Z\eta = 100Z_\zeta$ and $Z_\zeta = j \times 10^{-4}$.
N.B.—When $\psi = 0$, $Z_e = 1 \times 10^{-4}$.

In the case of an H$_0$-wave the guide will behave as if it had a uniform surface impedance given by eqn. (26). Consequently the graph of Z_e as a function of ψ is the same as for E-waves but with the abscissa in terms of (90° — ψ) instead of ψ.

All higher-order H-waves are unstable (as has been shown in Section 3.2), but the particular combination of the waves given by eqns. (36) or (38) can proceed down the guide without change of form and it represents, as explained, a spinning H-wave. Thus, if the wave guided by the structure is an H$_{11}$-wave, we get from eqn. (41) that

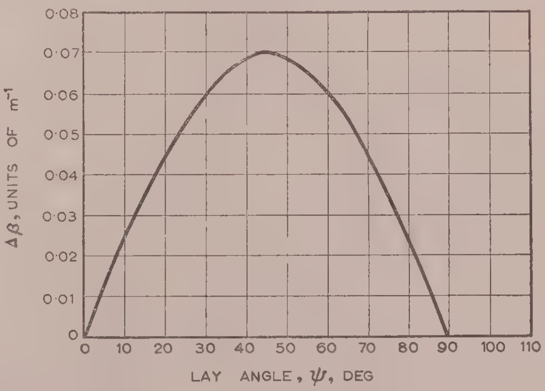


Fig. 7.—Wave spin coefficient $\Delta\beta$ of the H$_{11}$-wave as a function of the lay angle ψ.
Constants of the guide:
$f_0 = 24\,000$ Mc/s
$s = 2$cm
$Z\eta = 100Z_\zeta$
$Z_\zeta = j \times 10^{-4}$

$$\Delta\beta = 1{\cdot}4 \times 10^{-3}\frac{\tan\psi}{1+\tan^2\psi}\ \text{metres}^{-1} \quad . \quad . \quad (42)$$

The quantity $\Delta\beta$ is shown plotted in Fig. 7 as a function of ψ. It will be observed that the maximum wave spin occurs when

$\psi = 45^\circ$ and for this angle the lay of the ζ-helix is $0 \cdot 126$m, but the lay of the spinning wave is 90m.

(5) CONCLUSIONS

A complete analysis of wave propagation in circular anisotropic waveguides has been carried out. General formulae have been derived and these relate propagation and attenuation coefficients of the wave to the dimensions of the guide structure and their impedance components.

It has been shown that all E-waves and H_0-waves are stable in anisotropic guides and consequently in the case of these waves, so far as the behaviour of the wave is concerned, the anisotropic guide wall could be regarded as a uniform and homogeneous wall of surface impedance Z_e (which can be calculated). The anisotropic nature of the wall is of no consequence except that it introduces a surface impedance Z_e whose value depends on the lay angle ψ.

All higher-order H-waves have been shown to be unstable in anisotropic guides; but a certain combination of H-waves (of the same order) given by eqn. (38) can proceed down an anisotropic guide without change of form. The stable solution is a "spinning H-wave" and this was shown to have the same sense of spin as that of the helix of lesser surface impedance. The lay of the spinning H-wave, however, is much longer than the lay of the helix of the principal surface impedance.

(6) ACKNOWLEDGMENTS

The author is indebted to Mr. L. Lewin, who has read the manuscript and made interesting comments.

Acknowledgment is made also to Standard Telecommunication Laboratories Ltd., for facilities granted in the preparation of the manuscript and permission to publish the paper.

(7) REFERENCES

(1) KARBOWIAK, A. E.: "Theory of Imperfect Waveguides: the Effect of Wall Impedance," *Proceedings I.E.E.*, Paper No. 1841 R, September, 1955 (**102 B**, p. 698).

(2) KARBOWIAK, A. E.: "The E–H Surface Wave," *Wireless Engineer*, 1954, **31**, p. 71.

(3) SENSIPER, S.: "Electromagnetic Wave Propagation on Helical Structures (A Review of Recent Progress)," *Proceedings of the Institute of Radio Engineers*, 1955, **43**, p. 149.

(8) APPENDIX

(8.1) A Note on the Stability of a Wave Mode in a Waveguide

The meaning of the term "stability" as applied to a mode in a waveguide has been explained by the author elsewhere,[1] and here some additional aspects of the concept of stability will be explained, without the recourse to formal mathematical treatment.

Any wave mode [e.g. that derived from eqn. (6) or eqn. (19)] is a mathematical possibility in a perfectly conducting waveguide whose geometry is perfect, in the sense that the waveguide inner wall forms a right cylinder of constant cross-section. The cylinder cross-section can be a rectangle, a circle, an ellipse, etc. However, so far as any practical applications are concerned, a waveguide can neither be perfectly conducting nor be endowed with an impeccable geometry.

We may find, for example, that owing to a small distortion of the originally perfect waveguide a wave mode will split into two or more component waves travelling with different phase velocities. In addition, the wave pattern of the associated wave will continuously change in form. Such a wave mode is referred to as being unstable with respect to the particular guide perturbation. There are many examples of unstable modes: all higher-order E- and H-modes are unstable in imperfectly conducting rectangular waveguides, the H_{01}- and E_{11}-modes are unstable in a curved circular waveguide, etc.

The appropriate mathematical tool for dealing with imperfect waveguides and investigating the stability of waveguide modes is the perturbation calculus, and here the order of the quantities involved is of importance. For example, if the perfectly conducting walls of a waveguide are replaced by an impedance sheet of a surface impedance, Z_s (such that $Z_s \ll 1$), then, unless the change in some of the propagation coefficients or other field parameters is of the order of Z_s, or larger, Z_s is said to have a negligible influence on the field of the guided mode. In particular, in the case of E-waves, discussed in Section 3.1, it has been shown that, to the order of Z_η and Z_ζ quantities, all waves are stable, but that the propagation coefficients are changed by amounts of the order of Z_η and Zr. On the other hand, it has been shown in the case of higher-order H-waves, that, to the order of Z_η and Z_ζ quantities, there are two distinct values of the propagation coefficient. In other words the modes are unstable.

(8.2) The Stable Solution for H-Waves

Using the coupling coefficients, eqn. (34), we find that the stable solutions are

$$\left.\begin{aligned} & H_{z1} = J_m(hr) \exp(-jm\phi - \gamma_1 z) \\ \text{and} \qquad & H_{z2} = J_m(hr) \exp(jm\phi - \gamma_2 z) \end{aligned}\right\} \quad . \quad . \quad . \quad (43)$$

These are evidently left- and right-hand circularly polarized waves travelling at phase velocities γ_1 and γ_2 respectively.

The modes H_{z1} and H_{z2} can be looked upon as normal modes of the anisotropic region, and we may combine them to satisfy any launching requirements. For example, if we specify that the mode is to resemble the H_{mn}-mode at any particular point in the waveguide so that the mode can be matched to an isotropic waveguide, then we find that H_{z1} and H_{z2} must be combined in equal proportion. The combination solution [eqn. (35)] is a mode peculiar to an anisotropic waveguide and is more convenient than the solutions given by eqn. (43). This single wave function represents, as explained in Section 3.2, a spinning H-wave.

621.372.8: 621.3.011.21

Monograph No. 148 R
Aug. 1955

JUNCTION ADMITTANCE BETWEEN WAVEGUIDES OF ARBITRARY CROSS-SECTIONS

By E. D. FARMER, B.A.

(*The paper was first received 4th January, and in revised form 3rd May*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in August*, 1955.

SUMMARY

A general formula for the admittance of the junction between two waveguides of arbitrary and different cross-sections, coupled end-to-end by an aperture of arbitrary shape, is derived by the application of Schwinger's variational procedure. It is shown that, if the dominant modes of either waveguide have similar patterns over the coupling aperture, the junction may be represented approximately by a 2-terminal network. A general and simple definition of characteristic impedance is introduced which enables us to regard the junction as an "impedance mismatch" together with a "junction effect" owing to shunt susceptance. The restrictions required for an exact 2-terminal description are discussed. Simplified formulae, applicable when the waveguides have similar cross-sections, are derived. The symmetrical junction is also considered. The approximate 2-terminal theory is applied to the junction between two rectangular guides of different E-plane dimensions and the results obtained are compared with those derived elsewhere by a more rigorous method. In this way some idea of the accuracy of the theory and the limits of its applicability is obtained. The 2-terminal theory is also applied to a circular-to-rectangular transition, and the results are shown to be in favourable agreement with experiment. The behaviour of a waveguide of hexagonal cross-section is analysed. Finally, various aspects of the impedance definition are discussed.

(1) INTRODUCTION

A common form of discontinuity encountered in microwave circuits is the junction between two waveguides of different cross-sections coupled end-to-end. If such a junction lies entirely within a common transverse plane and is perfectly conducting except over the coupling aperture, it is possible to describe its behaviour in purely general terms. By deriving a general expression for the admittance between two waveguides of arbitrary and different cross-sections coupled by an aperture of arbitrary shape, we have the key to a whole class of waveguide problems. In particular, all problems dealing with transverse conducting diaphragms or with junctions between misaligned waveguides fall into this class.

Although the equivalent circuit of the junction requires, in general, a 3-terminal network, an approximate 2-terminal representation usually suffices provided that the characteristic impedances of the waveguides are defined in an appropriate manner. The restriction required for a 2-terminal description to be approximately valid is that the dominant modes of the two waveguides shall have similar patterns over the coupling aperture. This restriction is not a stringent one, since it will always be satisfied provided that the singularity at the junction is not extremely severe. If this condition is satisfied and if the characteristic impedances are so defined that their ratio takes a special form, the junction may be approximately represented by a 2-terminal network. As a consequence, there exists an optimum definition of characteristic impedance such that the singularity at the junction may be represented by a shunt susceptance (the so-called junction effect) together with a "change in characteristic impedance."

If more stringent conditions than those considered are imposed, special cases of a simpler nature arise. Thus, if the waveguide cross-sections and the aperture differ only slightly in shape and area, the shunt susceptance or "junction effect" becomes negligible. In such a case, the singularity of the junction may be regarded as being entirely due to a change in characteristic impedance. Junctions of this type are particularly interesting from the theoretical point of view, as the determination of the circuit parameters involves very little computation. Another case of interest occurs when the dominant mode patterns are identical over the aperture, for then the 2-terminal equivalent-circuit representation is perfectly rigorous. Into this class falls the junction between two rectangular waveguides of different E-plane but identical H-plane dimensions, together with all symmetrical junctions.

(2) THE WAVE FUNCTIONS

Consider the end-to-end junction between two arbitrary cylindrical waveguides illustrated in Fig. 1. It is assumed that both waveguides support only one propagating mode and that the dominant mode in either waveguide is an H-type mode. It is further assumed that both waveguides enclose a common medium, although the results are easily generalized to include the case of two waveguides containing different media.

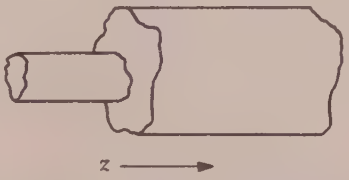


Fig. 1.—Junction between two arbitrary waveguides.

If a unit-amplitude dominant mode is incident in guide 1, the singularity at the junction will, in general, give rise to a reflected dominant mode of amplitude R (where R is the reflection coefficient) together with a double infinitude of non-propagating H_{mn} and E_{mn} modes.

In such a case the longitudinal component of the magnetic field will be of the form

$$H_z = (\varepsilon^{-j\beta z} + R\varepsilon^{+j\beta z})\psi + \sum_{m,n}\psi_{mn}\varepsilon^{\gamma_{mn}z} \quad . \quad . \quad (1)$$

where $j\beta$ and γ_{mn} are the propagation coefficients of the dominant and non-propagating H modes respectively and where ψ and ψ_{mn} are the wave functions associated with these modes. ψ and ψ_{mn} are functions of the transverse co-ordinates only and satisfy the wave equations

$$\left.\begin{array}{l}(\text{divgrad} + k^2)\psi = 0 \quad (\text{divgrad} + k_{mn}^2)\psi_{mn} = 0 \\ \text{together with the boundary conditions} \\ \dfrac{\partial\psi}{\partial v} = \dfrac{\partial\psi_{mn}}{\partial v} = 0 \quad \text{on the conductor}\end{array}\right\} \quad . \quad (2)$$

Correspondence on Monographs is invited for consideration with a view to publication.

The paper is a communication from the Staff of the Research Laboratories of The General Electric Company, Limited, Stanmore, Middlesex, England.

$\partial/\partial v$ denotes differentiation along the normal; k and k_{mn} are the transverse wave numbers of the dominant mode and H_{mn} mode respectively; they satisfy the equations

$$k^2 = \omega^2\mu\epsilon - \beta^2; \quad k_{mn}^2 = \omega^2\mu\epsilon + \gamma_{mn}^2 \quad . \quad . \quad (3)$$

where ω, μ and ϵ are the angular frequency, permeability and permittivity respectively. (Rationalized M.K.S. units are employed throughout.)

Similarly, the longitudinal component of the electric field will take the form

$$E_z = \sum_{m,n} B_{mn}\varepsilon^{\rho_{mn}z}\phi_{mn} \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where ϕ_{mn} is the transverse wave function of the E_{mn} mode satisfying the wave equation

$$\left.\begin{array}{l} (\mathrm{divgrad} + \sigma_{mn}^2)\phi_{mn} = 0 \\ \text{and the boundary condition} \\ \phi_{mn} = 0 \text{ on the conductor} \end{array}\right\} \quad . \quad . \quad . \quad . \quad (5)$$

ρ_{mn} and σ_{mn} are related by the equation

$$\sigma_{mn}^2 = \omega^2\mu\epsilon + \rho_{mn}^2 \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The operators "div" and "grad" are transverse operators, i.e. they act on the transverse co-ordinate only.

The functions ψ, ψ_{mn} and ϕ_{mn} are determined by the geometry of the waveguide cross-section; they are real, frequency independent, and eigenfunctions of eqns. (2) and (5). The wave numbers k, k_{mn} and σ_{mn} are also characteristic of the waveguide cross-section only; they are the corresponding eigenvalues of the operator "divgrad."

Furthermore, owing to the self-adjoint nature of the "divgrad" operator the eigenfunctions (ψ, ψ_{mn}) and ϕ_{mn} are orthogonal. They may be normalized and chosen so that

$$\int \psi_{mn}^2 dS = \int \phi_{mn}^2 dS = 1$$

where the integral extends over the waveguide cross-section. The dominant-mode eigenfunction will not be considered as normalized for reasons of convenience. In this case the orthogonality relations may then be summarized in the equations

$$\left.\begin{array}{l} \int \psi\psi_{mn}dS = \int \mathrm{grad}\,\psi \;\; \mathrm{grad}\,\psi_{mn}dS = 0 \\ k_{mn}^2 \int \psi_{mn}\psi_{m'n'}dS = \int \mathrm{grad}\,\psi_{mn}\,\mathrm{grad}\,\psi_{m'n'}dS = k_{mn}^2\delta_{mm'}\delta_{nn'} \\ \sigma_{mn}^2 \int \phi_{mn}\phi_{m'n'}dS = \int \mathrm{grad}\,\phi_{mn}\,\mathrm{grad}\,\phi_{m'n'}dS = \sigma_{mn}^2\delta_{mm'}\delta_{nn'} \\ \int \bar{u}(\mathrm{grad}\,\phi_{mn} \times \mathrm{grad}\,\psi_{m'n'})dS = 0 \end{array}\right\} \quad (7)$$

where $\delta_{mm'}$ is the Kronecker delta $\left(\delta_{mm'} = \begin{array}{ll} 1 & m = m' \\ 0 & m \neq m' \end{array}\right)$ and $\boldsymbol{u}$ is a unit sector in the direction of the z-axis.

It follows further from the monochromatic Maxwell equations, curl $H = j\omega\epsilon E$, curl $E = -j\omega\mu H$, that the transverse field components at the junction plane ($z = 0$) take the form

$$\left.\begin{array}{l} \boldsymbol{E}_t \times \boldsymbol{u} = \dfrac{j\omega\mu}{k^2}(1 + R)\,\mathrm{grad}\,\psi + j\omega\mu \displaystyle\sum_{m,n} \frac{A_{mn}}{k_{mn}^2}\,\mathrm{grad}\,\psi_{mn} \\ \qquad\qquad - \boldsymbol{u} \times \displaystyle\sum_{m,n} \frac{\rho_{mn}B_{mn}}{\sigma_{mn}^2}\,\mathrm{grad}\,\phi_{mn} \\ \boldsymbol{H}_t = -\dfrac{j\beta}{k^2}(1 - R)\,\mathrm{grad}\,\psi + \displaystyle\sum_{m,n} \frac{A_{mn}\gamma_{mn}}{k_{mn}^2}\,\mathrm{grad}\,\psi_{mn} \\ \qquad\qquad - j\omega\epsilon\boldsymbol{u} \times \displaystyle\sum_{m,n} \frac{B_{mn}}{\sigma_{mn}^2}\,\mathrm{grad}\,\phi_{mn} \end{array}\right\} \quad (8)$$

(3) VARIATIONAL FORM OF THE ADMITTANCE

It follows from the orthogonality relations, eqns. (7), and from the expression for $\boldsymbol{E}_t \times \boldsymbol{u}$ in eqn. (8) that the amplitudes, $1 + R$, A_{mn} and B_{mn}, are given by

$$\left.\begin{array}{l} 1 + R = \dfrac{jk^2}{\omega\mu}\,\dfrac{\int \boldsymbol{u} \times \boldsymbol{E}_t\,\mathrm{grad}\,\psi dS}{\int (\mathrm{grad}\,\psi)^2\,dS} \\ A_{mn} = j/\omega\mu \int \boldsymbol{u} \times \boldsymbol{E}_t\,\mathrm{grad}\,\psi_{mn}dS, \\ B_{mn} = 1/\rho_{mn} \int \boldsymbol{E}_t\,\mathrm{grad}\,\phi_{mn}dS \end{array}\right\} \quad . \quad . \quad (9)$$

These expressions may be used to eliminate the amplitude in eqn. (8) for $\bar{H}_t$. If this is done and the resulting equation is scalar multiplied by $\bar{u} \times \bar{E}_t$ and integrated over the cross-section we obtain

$$\left.\begin{array}{l} \int \boldsymbol{u}\boldsymbol{E}_t \times \boldsymbol{H}_t dS = \beta/\omega\mu \dfrac{(\int \boldsymbol{u} \times \boldsymbol{E}_t\,\mathrm{grad}\,\psi dS)^2}{\int (\mathrm{grad}\,\psi)^2\,dS}\left(\dfrac{1 - R}{1 + R}\right) \\ \qquad + \displaystyle\sum_{m,n} \frac{j\gamma_{mn}}{\omega\mu k_{mn}^2}(\int \boldsymbol{u} \times \boldsymbol{E}_t\,\mathrm{grad}\,\psi_{mn}dS)^2 \\ \qquad - \displaystyle\sum_{m,n} \frac{j\omega\epsilon}{\sigma_{mn}^2\rho_{mn}}(\int \boldsymbol{E}_t\,\mathrm{grad}\,\phi_{mn}dS)^2 \end{array}\right\} \quad . \quad (10)$$

However, $\boldsymbol{E}_t$ is zero over those parts of the junction plane which do not coincide with the aperture A, and consequently integrals involving $\boldsymbol{E}_t$ may be considered as being taken over the aperture A only. Thus we have

$$\int \boldsymbol{u}\boldsymbol{E}_t \times \boldsymbol{H}_t dS = \int \boldsymbol{u}\boldsymbol{E}_t \times \boldsymbol{H}_t dA$$

Thus the left-hand side of eqn. (10) is equal to the instantaneous power flow across the aperture.

However, if ψ', ψ'_{mn}, ϕ'_{mn} are the corresponding wave functions on the other side of the aperture in waveguide 2, the instantaneous power flow for the same field [eqns. (1) and (4)] at the aperture may be expanded as an analogous expression in ψ', ψ'_{mn} and ϕ'_{mn}. If such an expansion is effected and the two expressions for the power flow are equated the resulting expression may be solved for the relative junction admittance $(1 - R)/(1 + R)$ to give the result

$$\begin{array}{l} \dfrac{1 - R}{1 + R} = \dfrac{\beta'}{\beta}\,\dfrac{\int (\mathrm{grad}\,\psi)^2 dS}{\int (\mathrm{grad}\,\psi')^2 dS'}\left(\dfrac{\int \boldsymbol{u} \times \boldsymbol{E}_t\,\mathrm{grad}\,\psi' dA}{\int \boldsymbol{u} \times \boldsymbol{E}_t\,\mathrm{grad}\,\psi dA}\right)^2 \\ \qquad - \displaystyle\sum_{m,n} \frac{j\gamma_{mn}}{\beta k_{mn}^2} \int (\mathrm{grad}\,\psi)^2\,dS\left(\frac{\int \boldsymbol{u} \times \boldsymbol{E}_t\,\mathrm{grad}\,\psi_{mn}dA}{\int \boldsymbol{u} \times \boldsymbol{E}_t\,\mathrm{grad}\,\psi dA}\right)^2 \\ \qquad - \displaystyle\sum_{m,n} \frac{j\gamma'_{mn}}{\beta k'^2_{mn}} \int (\mathrm{grad}\,\psi)^2\,dS\left(\frac{\int \boldsymbol{u} \times \boldsymbol{E}_t\,\mathrm{grad}\,\psi'_{mn}dA}{\int \boldsymbol{u} \times \boldsymbol{E}_t\,\mathrm{grad}\,\psi dA}\right)^2 \\ \qquad + \displaystyle\sum_{m,n} \frac{j\omega^2\mu\epsilon}{\rho_{mn}\sigma_{mn}^2\beta} \int (\mathrm{grad}\,\psi)^2 dS\left(\frac{\int \boldsymbol{E}_t\,\mathrm{grad}\,\phi_{mn}dA}{\int \boldsymbol{u} \times \boldsymbol{E}_t\,\mathrm{grad}\,\psi dA}\right)^2 \\ \qquad + \displaystyle\sum_{m,n} \frac{j\omega^2\mu\epsilon}{\beta\rho'_{mn}\sigma'^2_{mn}} \int (\mathrm{grad}\,\psi)^2 dS\left(\frac{\int \boldsymbol{E}_t\,\mathrm{grad}\,\phi'_{mn}dA}{\int \boldsymbol{u} \times \boldsymbol{E}_t\,\mathrm{grad}\,\psi dA}\right)^2 . \quad (11) \end{array}$$

where S and S' are the respective waveguide cross-sectional areas and A is the aperture area.

It can be shown [cf. References (1) and (2)] that expressions of the type on the right-hand side of eqn. (11) are stationary with respect to variations of the aperture field about its correct value.

Such an expression may therefore be submitted to the so-called Rayleigh–Ritz procedure in which the aperture field is expanded in terms of a minimal sequence of functions such as the normal modes on one side of the junction. The unknown amplitudes C_n (say) may then be determined by the stationary conditions

$$\frac{\partial}{\partial C_n}\left(\frac{1 - R}{1 + R}\right) = 0$$

The accurate aperture field so obtained is then substituted into he expression for the admittance, and a highly accurate dmittance formula is obtained.

However, the expressions resulting from such a procedure are xceedingly cumbersome, and it will suffice to take the incident node as a first-order approximation to the aperture field.)wing to the stationary nature of eqn. (11), this result is a econd-order approximation for the admittance itself.

(4) THE APPROXIMATE FORM OF THE ADMITTANCE

If the incident dominant mode is substituted as an approxination to the aperture field, i.e. if we put $\boldsymbol{E}_t = \boldsymbol{u} \times \text{grad}\,\psi$ in qn. (11), the approximation obtained for the admittance is

$$\frac{1-R}{1+R} = P + jQ \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

where

$$P = \frac{\beta'}{\beta}\,\frac{(\psi,\psi)_S}{(\psi',\psi')_{S'}}\left[\frac{(\psi,\psi')_A}{(\psi,\psi)_A}\right] \quad . \quad . \quad . \quad . \quad (13)$$

nd

$$\begin{aligned} Q = &\sum_{m,n}\frac{\omega^2\mu\epsilon}{\beta\sigma_{mn}^2\rho_{mn}}(\psi,\psi)_S\left[\frac{\int \boldsymbol{u}\,\text{grad}\,\psi \times \text{grad}\,\phi_{mn}dA}{(\psi,\psi)_A}\right]^2 \\ &+ \sum_{m,n}\frac{\omega^2\mu\epsilon}{\beta'\sigma_{mn}'^2\rho_{mn}'}(\psi,\psi)_S\left[\frac{\int \boldsymbol{u}\,\text{grad}\,\psi \times \text{grad}\,\phi_{mn}'dA}{(\psi,\psi)_A}\right]^2 \\ &- \sum_{m,n}\frac{\gamma_{mn}}{\beta k_{mn}^2}(\psi,\psi)_S\left[\frac{\int \text{grad}\,\psi\,\text{grad}\,\psi_{mn}dA}{(\psi,\psi)_A}\right]^2 \\ &- \sum_{m,n}\frac{\gamma_{mn}'}{\beta' k_{mn}'^2}(\psi,\psi)_S\left[\frac{\int \text{grad}\,\psi\,\text{grad}\,\psi_{mn}'dA}{(\psi,\psi)_A}\right]^2 \quad . \quad . \quad (14) \end{aligned}$$

where we have used the notation $(\psi,\psi')_A$ to denote the inner roduct of the gradients of ψ and ψ' over A, i.e.

$$(\psi,\psi')_A \equiv \int \text{grad}\,\psi\,\text{grad}\,\psi' dA \quad . \quad . \quad . \quad (15)$$

f R' denotes the reflection coefficient on the other side of the unction, then similarly

$$\frac{1-R'}{1+R'} = P' + jQ' \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where

$$P' = \frac{\beta}{\beta'}\,\frac{(\psi',\psi')_{S'}}{(\psi,\psi)_S}\left[\frac{(\psi,\psi')_A}{(\psi',\psi')_A}\right]^2 \quad . \quad . \quad . \quad (17)$$

nd where Q' is the expression which is obtained from Q by nterchanging the primes in eqn. (14).

On account of the lossless condition the reflection coefficients R and R' determine the behaviour of the junction completely including its transfer properties except in the trivial case $R = R' = 0$). The values of R and R' are given quite generally y eqns. (12) and (16) for the type of junction under conideration.

However, when we attempt to use the expressions for R and R' n order to set up a 3-terminal equivalent circuit, the task becomes rohibitively complicated. We therefore seek approximations vhich will lead to simpler descriptions of the behaviour of the unction.

(5) THE APPROXIMATE 2-TERMINAL REPRESENTATION

Before discussing the validity of a 2-terminal equivalent-circuit epresentation of the junction it is necessary to make some reliminary analytical considerations.

We obtain from eqns. (13) and (17)

$$\sqrt{PP'} = \frac{(\psi,\psi')_A^2}{(\psi,\psi)_A(\psi',\psi')_A} \quad . \quad . \quad . \quad . \quad (18)$$

Two facts may be noted about this expression. In the first place, by virtue of Schwarz's inequality, we have

$$(\psi,\psi)_A(\psi',\psi')_A \geqslant (\psi,\psi')_A^2$$

Hence

$$\sqrt{PP'} \leqslant 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

the equality sign holds when the functions ψ and ψ' are linearly dependent over the aperture, i.e. when $\psi = \alpha\psi'$, where α is a constant.

The measure of the linear dependence of ψ and ψ' is given by the lower eigenvalue (which is always non-negative) of the Gram matrix Γ where

$$\Gamma = \begin{bmatrix} (\psi,\psi)_A & (\psi,\psi')_A \\ (\psi',\psi)_A & (\psi',\psi')_A \end{bmatrix}$$

If this smaller eigenvalue is much smaller than $(\psi,\psi)_A^2$, the functions may be said to be "almost linearly dependent," and consequently we have

$$\sqrt{PP'} \simeq 1$$

Similarly, if a constant value of α can be found such that $\psi' = \alpha\psi + \delta$, where δ is a small function in the sense that $(\psi,\psi)_A(\delta,\delta)_A - (\psi,\delta)_A^2$ is much smaller than $(\psi,\psi')_A^2$, then

$$\sqrt{PP'} = 1 - 0(\delta^2) \quad . \quad . \quad . \quad . \quad . \quad (20)$$

If undefined quantities Y_0 and Y_0' are introduced such that their ratio satisfies the relation

$$\frac{Y_0'}{Y_0} = \sqrt{\frac{P}{P'}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

the inequality (19) gives

$$P \leqslant \frac{Y_0'}{Y_0} \leqslant \frac{1}{P'} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

Furthermore, if ψ and ψ' are almost linearly dependent in the above sense, then from eqn. (20)

$$P \simeq \frac{Y_0'}{Y_0} \qquad P' \simeq \frac{Y_0}{Y_0'}$$

In addition, under these conditions, it may be shown that

$$Q'Y_0' \simeq QY_0 = B\,(\text{say})$$

Thus eqns. (12) and (16) take the approximate forms

$$\frac{1-R}{1+R} = \frac{Y_0'}{Y_0} + j\frac{B}{Y_0} \quad \frac{1-R'}{1+R'} = \frac{Y_0}{Y_0'} + j\frac{B}{Y_0'} \quad . \quad (23)$$

However, these equations are just those which define the reflection coefficients at a 2-terminal junction of susceptance B, between two waveguides of characteristic admittances Y_0 and Y_0'.

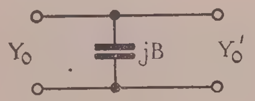


Fig. 2.—Equivalent circuit.

Thus the approximate equations are completely equivalent to those arising from the 2-terminal equivalent circuit illustrated in Fig. 2.

The relative susceptance is given by $B/Y_0 = Q$, where Q is defined by eqn. (14). Alternatively, we may put $B/Y_0' = Q'$ and so define B/Y_0 as $Q'Y_0/Y_0'$. The ratio of the admittances is determined by eqns. (21), (13) and (17):

$$\frac{Y_0'}{Y_0} = \frac{\lambda_g}{\lambda_g'}\frac{\int(\text{grad }\psi)^2dS}{\int(\text{grad }\psi')^2dS'}\frac{\int(\text{grad }\psi')^2dA}{\int(\text{grad }\psi)^2dA} \quad . \quad . \quad (25)$$

where $\lambda_g = 2\pi/\beta$ and $\lambda_g' = 2\pi/\beta'$ are the respective guide wavelengths.

It is evident that, if the dominant modes in either waveguide have similar patterns over the coupling aperture, the above 2-terminal representation of the junction is accurate to a second order (when the dominant-mode aperture patterns differ by a first-order small function). The 2-terminal representation given will always give a fairly accurate description of the junction provided that the singularity at the junction plane is not extremely severe.

(6) THE CHARACTERISTIC IMPEDANCES

Let us consider the nature of the characteristic impedances Z_0 and Z_0' of the two waveguides forming the junction. As is well known, the concept of impedance as applied to waveguides has an arbitrary element which is decided by convention and which is immaterial so long as we are concerned with the constructs within a guide of a single uniform cross-section, since then only relative impedances are relevant. Such conventional definitions of characteristic impedance Z_0 all take the form

$$Z_0 = M\frac{\lambda_g}{\lambda}\sqrt{\frac{\mu}{\epsilon}}\text{ ohms} \quad . \quad . \quad . \quad . \quad . \quad (26)$$

where M is a dimensionless frequency-independent constant; the differing conventions merely assign different values to M.

When two waveguides are coupled together, these connections are usually incompatible with the further requirement for a useful impedance concept, namely that the ratio of impedances should account for the "impedance mismatch" part of the reflection coefficients. It is possible to introduce a definition of characteristic impedance which does satisfy this requirement. To do this, it is necessary to make a slight generalization of eqn. (26) in that M, instead of being a constant characteristic of the waveguide cross-section only, is regarded as being dependent on the geometry of the aperture A through which the waveguide is coupled to a second one.

It is then permissible to define the constant M in accordance with the equation

$$M = \frac{\int(\text{grad }\psi)^2dS}{\int(\text{grad }\psi)^2dA} = \frac{\int E_d^2dS}{\int E_d^2dA} \quad . \quad . \quad . \quad (27)$$

where E_d is the transverse component of the electric field associated with the dominant mode of the waveguide. The expression on the right-hand side of eqn. (27) is quite independent of the amplitude of E_d (and its dependence on the longitudinal coordinate z).

With the definition of M, the characteristic impedance Z_0 becomes

$$Z_0 = \frac{\lambda_g}{\lambda}\sqrt{\frac{\mu}{\epsilon}}\frac{\int E_d^2dS}{\int E_d^2dA} \quad . \quad . \quad . \quad . \quad (28)$$

The characteristic impedance of the second waveguide is similarly

$$Z_0' = \frac{\lambda_g'}{\lambda}\sqrt{\frac{\mu}{\epsilon}}\frac{\int E_d'^2dS'}{\int E_d'^2dA} \quad . \quad . \quad . \quad . \quad (29)$$

and their ratio is

$$\frac{Z_0'}{Z_0} = \frac{\lambda_g'}{\lambda_g}\frac{\int E_d^2dS'}{\int E_d^2dS}\frac{\int E_d^2dA}{\int E_d'^2dA} \quad . \quad . \quad . \quad . \quad (30)$$

This equation is completely equivalent to eqn. (25), which defines the admittance ratio. It follows that the definitions given in eqns. (28) and (29) of the characteristic impedance are just those which enable us to regard the junction as an "impedance mismatch" together with a shunt effect. Indeed, in view of the inequalities of eqn. (22), which may be written

$$P \leqslant \frac{Z_0}{Z_0'} \leqslant \frac{1}{P'} \quad . \quad (25)$$

it is evident that, even when the aperture field patterns differ considerably, the definition [eqn. (30)] of the ratio of the impedances represents a kind of mean value which accounts for the impedance-mismatch contributions to the reflection coefficients on either side of the junction. Although, in the case of a particular junction, it may be possible to obtain more accurate 2-terminal descriptions by moving the reference planes, the definition [eqn. (28)] is, in view of its generality and simplicity, a formula of some practical significance.

By carrying out the analysis in a more general manner it is easily shown that if the two waveguides enclose different media whose permittivities and permeabilities are μ, ϵ and μ', ϵ', respectively, the corresponding formulae for the characteristic impedances become

$$\left.\begin{aligned} Z_0 &= \omega\mu\lambda_g\frac{\int E_d^2dS}{\int E_d^2dA} \\ Z_0' &= \omega\mu'\lambda_g'\frac{\int E_d'^2dS'}{\int E_d'^2dA} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (31)$$

Eqns. (31) represent the final generalized forms of the characteristic impedances. Although they were initially derived for H modes, they are equally applicable to TEM modes. In addition, for a single uniform waveguide supporting two different media, the reflection coefficient at the interface between the media is given exactly by the impedance mismatch

$$\text{i.e.} \qquad \frac{1-R}{1+R} = \frac{Z_0}{Z_0'} \qquad \frac{1-R'}{1+R'} = \frac{Z_0'}{Z_0}$$

where Z_0 and Z_0' are given by eqn. (31).

Thus, even when the waveguides contain different media, the definition [eqn. (31)] of the characteristic impedance still represents an optimum one in that the type of junction under consideration may be regarded as an impedance mismatch together with a "junction effect" owing to the shunt susceptance B.

(7) PHYSICAL INTERPRETATION OF THE SUSCEPTANCE

The relative susceptance of the junction is given by

$$B/Y_0 = Q$$

where Q is given by eqn. (14).

The only frequency-dependent quantities involved in eqn. (14) are $\omega^2\mu\epsilon/\beta\rho_{mn}$ and γ_{mn}/β, and the corresponding primed quantities. However, by the use of eqns. (3) and (6) the following expansions may be obtained:

$$\left.\begin{aligned} \gamma_{mn} &= k_{mn}\left(1 - \frac{1}{2}\frac{\omega^2\mu\epsilon}{k_{mn}^2} + \ldots\right) \\ \frac{\omega^2\mu\epsilon}{\rho_{mn}} &= \frac{\omega^2\mu\epsilon}{\sigma_{mn}}\left(1 + \frac{1}{2}\frac{\omega^2\mu\epsilon}{\sigma_{mn}^2} - \ldots\right) \end{aligned}\right\} \quad . \quad . \quad (32)$$

These series converge because $\omega^2\mu\epsilon < k_{mn}^2$ and $\omega^2\mu\epsilon < \sigma_{mn}^2$ when there is only one propagating mode in the waveguide.

If also the definition of eqn. (28) is adopted for the characteristic impedance Z_0, by means of eqns. (14) and (32) we may obtain an expression for the susceptance B:

$$\left.\begin{aligned} B &= \omega C - \frac{1}{\omega L} \\ \text{where}\qquad C &= C_1 + C_2(\omega) \\ \text{and}\qquad \frac{1}{L} &= \frac{1}{L_1} - \frac{1}{L_2(\omega)} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (33)$$

$$\begin{aligned} \text{and}\quad C_1 &= K\left[\int (\text{grad}\,\psi)^2 dA\right]^{-1} \sum_{m,n} (P_{mn} + P'_{mn}) \\ \frac{1}{L_1} &= \mu\left[\int (\text{grad}\,\psi)^2 dA\right]^{-1} \sum_{m,n} (Q_{mn} + Q'_{mn}) \end{aligned} \quad . \quad (34)$$

The quantities P_{mn}, P'_{mn}, Q_{mn}, and Q'_{mn} are given by

$$\left.\begin{aligned} P_{mn} &= \frac{(\int \bar{u}\,\text{grad}\,\psi \times \text{grad}\,\phi_{mn} dA)^2}{\sigma^3_{mn}} \\ P'_{mn} &= \frac{(\int \bar{u}\,\text{grad}\,\psi \times \text{grad}\,\phi'_{mn} dA)^2}{\sigma'^3_{mn}} \\ Q_{mn} &= \frac{(\int \text{grad}\,\psi\,\text{grad}\,\psi_{mn} dA)^2}{k_{mn}} \\ Q'_{mn} &= \frac{(\int \text{grad}\,\psi\,\text{grad}\,\psi'_{mn} dA)^2}{k'_{mn}} \end{aligned}\right\} \quad . \quad . \quad (35)$$

C_2 and $1/L_2$ are small frequency correction terms. If, as is usually sufficient, only the first two terms of expansion (32) are retained, these correction terms are given by

$$\left.\begin{aligned} C_2(\omega) &= \tfrac{1}{2}\omega^2\mu\epsilon^2\left[\int (\text{grad}\,\psi)^2 dA\right]^{-1} \sum_{m,n} \left(\frac{P_{mn}}{\sigma^2_{mn}} + \frac{P'_{mn}}{\sigma'^2_{mn}}\right) \\ \frac{1}{L_2(\omega)} &= \tfrac{1}{2}\omega^2\epsilon\left[\int (\text{grad}\,\psi)^2 dA\right]^{-1} \sum_{m,n} \left(\frac{Q_{mn}}{k^2_{mn}} + \frac{Q'_{mn}}{k'^2_{mn}}\right) \end{aligned}\right\} \quad . \quad (36)$$

The quantities C_1 and $1/L_1$ are the major contributors to C and $1/L$, and hence the susceptance B may be represented as a capacitance and inductance in parallel (see Fig. 3). In the interpretation of the equivalent circuit it must be borne in mind that both C and L vary slightly with frequency. If the contributions C_2 and $1/L_2$ are neglected, we obtain the so-called "quasi-static" solution for the susceptance. In any case, this susceptance may be regarded as a simple resonant circuit in the general case when both non-propagation H modes and E modes are present in the vicinity of the junction singularity. As is evident, the H modes give rise to the inductive part of the susceptance B whilst the E modes give rise to the capacitive part.

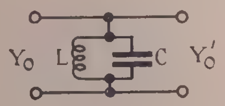


Fig. 3.—Equivalent circuit.

(8) EXACT 2-TERMINAL REPRESENTATION

If the dominant-mode patterns of either waveguide are identical over the coupling aperture, the 2-terminal representation of the junction is rigorous. This may be seen formally by the following argument. The transverse electric field in the junction plane vanishes except over the aperture, and hence the amplitude of the dominant mode in guide 1 is, from eqn. (9),

$$\frac{jk^2}{\omega\mu}\,\frac{\int \boldsymbol{u} \times \boldsymbol{E}_t\,\text{grad}\,\psi\, dA}{\int (\text{grad}\,\psi)^2 dS}$$

Similarly, the amplitude at the junction plane of the transmitted dominant mode is

$$\frac{jk'^2}{\omega\mu}\,\frac{\int \boldsymbol{u} \times \boldsymbol{E}_t\,\text{grad}\,\psi' dA}{\int (\text{grad}\,\psi')^2 dS}$$

Hence, if the functions ψ and ψ' are linearly dependent over A, i.e. if $\psi = \alpha\psi'$, where α is constant, and if one of the above amplitudes vanishes, so does the other; i.e. if a short-circuit is placed across one pair of terminals, a short-circuit exists across the other pair. Hence the impedance of the series arm in the equivalent network vanishes. The presence of a possible ideal transformer element may always be eliminated by a special choice of characteristic impedance, and furthermore, eqn. (30) defines the ratio of the impedance in precisely this manner. The equivalent circuit reduces to a shunt susceptance.

(9) THE COUPLING OF WAVEGUIDES OF SIMILAR CROSS-SECTIONS

If two waveguides have cross-sections which differ only slightly in shape and area and are coupled by an aperture which is similar to either cross-section, the dominant mode patterns are very nearly identical and the 2-terminal approximation is a very good one. Furthermore, in this case the shunt susceptance B is negligible, and consequently the junction may be regarded entirely as an impedance mismatch. The reflection coefficient is given approximately by the equation

$$\frac{1 - R}{1 + R} = \frac{Z_0}{Z'_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

Eqn. (30) for the ratio of the impedances simplifies to the form

$$\frac{Z_0}{Z'_0} = \frac{\lambda_g}{\lambda'_g}\,\frac{\int E^2_d dS}{\int E^2 dS'} \quad . \quad . \quad . \quad . \quad . \quad (38)$$

This formula is obtained by neglecting second-order small quantities in eqn. (30). It has the advantage that the relative impedances are determined (to this order of approximation) by the field pattern in one guide only, and a knowledge of both field patterns is not required. As a consequence, this formula is readily applied to those problems involving small perturbations of an original cross-section whose properties are known. In addition, it enables one to assess the manufacturing tolerances which are required to ensure that the standing-wave ratio of the junction is within certain specified limits.

(10) THE SYMMETRICAL JUNCTION

If the junction is symmetrical about the aperture plane, both waveguides are identical in cross-section and we have the case of a transversely-placed diaphragm or iris. The 2-terminal circuit is rigorous and the parameters simplify. Thus by putting

$$\psi' = \psi,\ \psi'_{mn} = \psi_{mn},\ \ \phi'_{mn} = \phi_{mn},\ \rho_{mn} = \gamma_{mn}$$

we obtain, from eqns. (14) and (24),

$$\begin{aligned} \frac{B}{Y_0} &= 2\sum_{m,n} \frac{\omega^2\mu\epsilon}{\beta k^2_{mn}\gamma_{mn}}(\psi, \psi)_S\left[\frac{\int u\,\text{grad}\,\psi \times \text{grad}\,\phi_{mn} dA}{(\psi, \psi)_A}\right]^2 \\ &\quad - 2\sum_{m,n} \frac{\gamma_{mn}}{\beta k^2_{mn}}(\psi, \psi)_S\left[\frac{\int \text{grad}\,\psi\,\text{grad}\,\psi_{mn}\, dA}{(\psi, \psi)_A}\right]^2 \end{aligned} \quad . \quad (39)$$

and, of course, we have $\quad Y'_0 = Y_0$

The susceptance B may be split into its inductive and capacitive parts, as has been shown previously.

(11) APPLICATION OF THE APPROXIMATE 2-TERMINAL THEORY TO THE JUNCTION BETWEEN TWO RECTANGULAR WAVEGUIDES OF DIFFERENT H-PLANE DIMENSIONS

As an application of the approximate 2-terminal theory, let us consider the junction between two rectangular waveguides of identical heights b, but different breadth dimensions a and a'.

It is evident that, when the dimensions a and a' differ considerably, the dominant-mode patterns are not at all similar over the aperture, and consequently the 2-terminal theory may be expected to break down under these conditions. This example is considered, not because it lends itself to a 2-terminal approximation, but because by comparing the results with a more rigorous solution it is possible to obtain some idea of the limitations in accuracy of the theory when it is applied under unfavourable conditions.

From symmetry considerations we see that the only non-propagating modes which arise in the vicinity of the junction are the H_{n0} modes with the odd n. The susceptance of the junction is therefore inductive (i.e. negative). To evaluate it we use the formula

$$\frac{B}{Y_0'} = Q', \text{ which is analogous to eqn. (24)}$$

Adopting the co-ordinate system of Fig. 4, we have

$$\operatorname{grad} \psi' = \boldsymbol{i}_x \cos \frac{\pi x}{a'} \qquad \operatorname{grad} \psi = \boldsymbol{i}_x \cos \frac{\pi x}{a}$$

$$\operatorname{grad} \psi_{n0}' = \frac{2}{a'} \boldsymbol{i}_x \cos \frac{n\pi x}{a'} \qquad \operatorname{grad} \psi_{n0} = \frac{2}{a} \boldsymbol{i}_x \omega s \frac{n\pi x}{a}$$

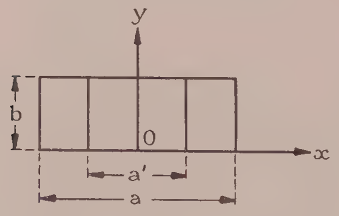


Fig. 4.—The junction plane.

Applying eqn. (25) we have

$$\frac{Y_0}{Y_0'} = \frac{\lambda_g'}{\lambda_g}\left(\alpha + \frac{1}{\pi}\sin \pi\alpha\right) \quad . \quad . \quad . \quad . \quad (40)$$

where $\alpha = a'/a$

To fix the ideas we consider the special case for which $\lambda = a$. Then we have

$$\frac{\lambda_g}{\lambda_g'} = \sqrt{\left(\frac{4 - \alpha^2}{3}\right)}$$

so that eqn. (40) becomes

$$\frac{Y_0}{Y_0'} = \sqrt{\left(\frac{3}{4 - \alpha^2}\right)}\left(\alpha + \frac{1}{\pi}\sin \pi\alpha\right) \quad . \quad . \quad . \quad (41)$$

(It is assumed throughout that $a' \leqslant a$.)

If we now use the formula

$$\frac{B}{Y_0} = \frac{B}{Y_0'}\frac{Y_0'}{Y_0} = Q'\frac{Y_0'}{Y_0}$$

and substitute the above wave functions into the expression for Q' we obtain

$$-\frac{B}{Y_0} = \frac{8}{\pi^2\sqrt{3}}\left(\alpha + \frac{1}{\pi}\sin \pi\alpha\right)^{-1} \sum_{n=3,5,\ldots} \frac{\alpha\sqrt{(n^2 - 4)}}{(n^2\alpha^2 - 1)^2}(1 - \cos n\pi\delta)$$

where $\delta = 1 - \alpha$

For large values of n, the term in n approaches

$$\frac{1}{n^3\alpha^3}(1 - \cos n\pi\delta)$$

and the series may be summed approximately and a correction term e computed. This gives

$$\left.\begin{aligned} -\frac{B}{Y_0} &= \frac{8}{\pi^2\sqrt{3}}\left(\alpha + \frac{1}{\pi}\sin \pi\alpha\right)^{-1} \\ &\qquad \left[5{\cdot}68\,\delta^2 \log_{10}\left(\frac{2{\cdot}84}{\delta}\right) + \cos \pi\delta + e - 1\right] \\ \text{where} \quad e &= 0{\cdot}0276(1 - \cos 3\pi\delta)[(1 - 1/9\alpha^2)^{-2} - 1{\cdot}34] \end{aligned}\right\} \quad (42)$$

Eqns. (41) and (42) give the solution of the problem in accordance with the approximate 2-terminal theory.

However, the present problem may be solved by a more rigorous procedure known as the equivalent-static method. In this method the aperture field is determined by going to the low-frequency limit. The fairly accurate field so obtained is then substituted into the stationary expression (11), and a good approximation to the junction admittance results. The results of this theory are quoted in Reference 3. As the required shift in the reference plane is of order of 10^{-3} times a wavelength, it may be neglected. The circuit parameters derived by this equivalent-static method, therefore, represent to a high degree of accuracy the "true" values with which the results of the approximate 2-terminal theory may be compared.

The results of the two theories are given graphically in Figs. 5 and 6. They are seen to agree well except in the neighbourhood of $a'/a = \frac{1}{2}$. The disagreement in this neighbourhood is due not so much to the approximations of the 2-terminal theory, as to the fact that, in deriving the reflection coefficients R and R', it was assumed that the incident dominant mode approximates to the aperture field pattern. This approximation is no longer a

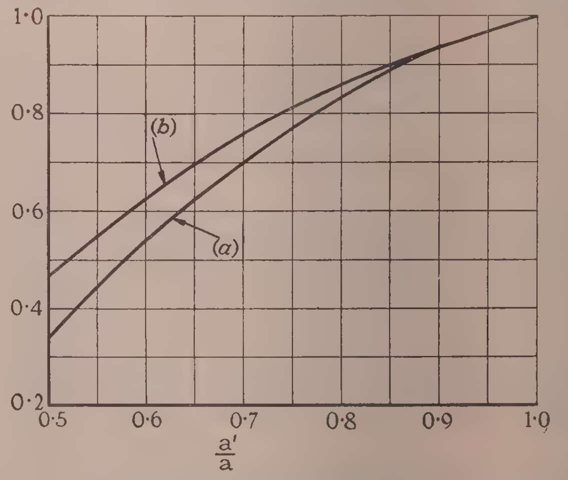


Fig. 5.—Curves of the admittance ratio.

(a) $(Y_0/Y_0')_1$ = Admittance ratio derived by equivalent static method.
(b) $Y_0'/Y_0)_2$ = Admittance ratio derived from general formula.

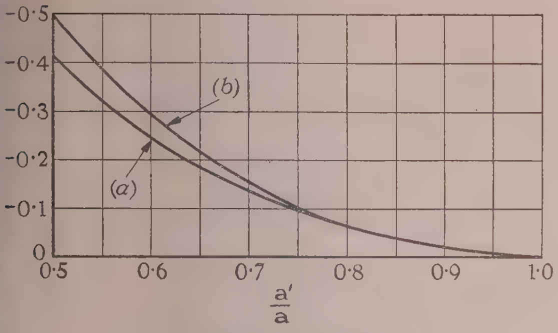


Fig. 6.—Curves of the susceptance.

(a) $(B/Y_0)_1$ = Relative susceptance according to equivalent static method.
(b) $(B/Y_0)_2$ = Relative susceptance derived from general formula.

good one, owing to the severe nature of the junction singularity when $a'/a = \frac{1}{2}$ and owing to the fact that when $a'/a = \frac{1}{2}$ the second waveguide is at the cut-off frequency. However, even under these unfavourable conditions, the ratio of the admittances is given to within 9%.

(12) APPLICATION OF THE 2-TERMINAL THEORY TO THE JUNCTION BETWEEN A RECTANGULAR AND A CIRCULAR WAVEGUIDE

The approximate 2-terminal theory may be applied to the junction between a circular and a rectangular waveguide coupled by an aperture, as illustrated in Fig. 7.

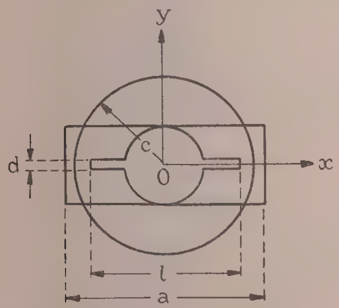


Fig. 7.—The junction plane.

The dominant-mode transverse components E_d and E'_d satisfy

$$E_d^2 = \cos^2\frac{\pi x}{a}, \quad E_d'^2 = J_0^2(kr) + J_2^2(kr) - 2J_0(kr)J_2(kr)\cos\theta$$

where $kc = 1\cdot842$ and where (r, θ) are polar co-ordinates.

If the characteristic impedances Z_0 and Z'_0 are defined according to eqns. (28) and (29), simple integrations give the following formulae for the characteristic admittances Y_0 and Y'_0:

$$Y_0 = \frac{\lambda}{\lambda_g}\sqrt{\frac{\epsilon}{\mu}}\left[\frac{ld}{ab} - \frac{d}{a} + \frac{d}{\pi b}\sin\frac{\pi b}{a} - \frac{d}{\pi b}\sin\frac{\pi l}{a} + \frac{\pi b}{4a} - \frac{1}{2}J_1\left(\frac{\pi b}{a}\right)\right]$$

$$Y'_0 = \frac{\lambda}{\lambda'_g}\sqrt{\frac{\epsilon}{\mu}}[(k^2c^2 - 1)J_0^2(kc)]^{-1}\left\{\frac{b^2}{8c^2}\left[J_0^2\left(\frac{kb}{2}\right) + J_1^2\left(\frac{kb}{2}\right)\right.\right.$$
$$\left.\left. + J_2^2\left(\frac{kb}{2}\right) - J_1\left(\frac{kb}{2}\right)J_3\left(\frac{kb}{2}\right)\right] + \frac{4d}{kc^2\pi}\int_{kb/2}^{kl/2} J_1'^2(x)dx\right\} \quad . \quad (43)$$

It has been found experimentally that, by adjusting the dimensions d and l, the susceptance B of the junction may be made to "resonate" out at a given frequency. Such a set of dimensions is given by

$$\frac{b}{a} = \frac{4}{9} \qquad \frac{c}{a} = 0\cdot486 \qquad \frac{d}{a} = \frac{1}{9} \qquad \frac{l}{a} = \frac{8}{9}$$

where λ is of the order $1\cdot4a$ for "resonance."

Under these conditions the "impedance mismatch" accounts entirely for the reflections at the junction. The substitution of the above values into eqn. (43) gives, for the admittance ratio,

$$\frac{Y'_0}{Y_0} = 0\cdot614\frac{\lambda_g}{\lambda'_g}$$

Y'_0/Y_0 is given graphically as a function of a/λ in Fig. 8; the results agree well with the experimentally determined values indicated.

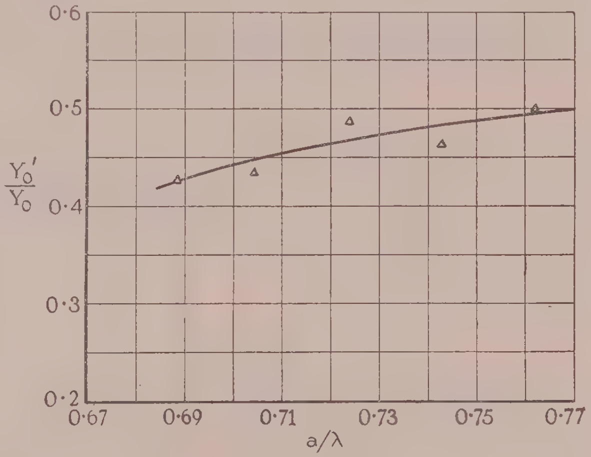


Fig. 8.—Curve of admittance ratio.

(Y'_0/Y_0) = Admittance ratio derived by means of eqns. (43).
Δ = Experimentally determined values.

(13) FURTHER APPLICATIONS AND DISCUSSION

It is evident that the simplest and most useful applications of the impedance formula will occur in those cases in which the susceptance B can be neglected. As indicated in Section 9, the junction between two similar waveguides is such a case, and the ratio of the impedances reduces to the form of eqn. (38), namely

$$\frac{Z_0}{Z'_0} = \frac{\lambda_g}{\lambda'_g}\frac{\int E_d^2 dS}{\int E_d'^2 dS'}$$

If the dominant mode pattern E_d and the guide wavelength λ_g are known for one of the guides only, the wavelength λ'_g in the second waveguide may be determined by simple perturbation theory. Consequently, the ratio of the impedances as given above may be determined without a knowledge of the dominant field pattern in the second guide.

This approach has been found to be both accurate and useful in many problems. In particular, the manufacture of die-cast waveguide is considerably facilitated by the introduction of a hexagonal cross-section, as shown in Fig. 9.

The "taper" angle α is regarded as small.

A simple perturbation calculation for the guide wavelength gives

$$\frac{\lambda_0}{\lambda'_g} = 1 - \alpha\frac{\lambda_0^2}{2\pi^2 a'b'} \quad . \quad . \quad . \quad . \quad . \quad (44)$$

where λ_0 is the guide wavelength when $\alpha = 0$.

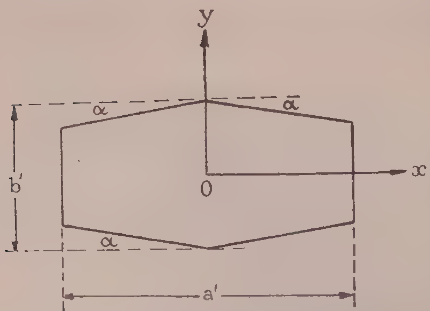


Fig. 9.—Hexagonal cross-section.

We consider the junction between this waveguide and a rectangular one of dimensions a and b.

From eqn. (44) we see that the guide wavelengths are equal if

$$a' = a\left(1 + \frac{2\alpha a}{\pi^2 b}\right) \quad . \quad . \quad . \quad . \quad . \quad (45)$$

In this case, by substituting the dominant mode pattern $E_d = \cos(\pi x/a)$ in eqn. (38), we obtain for the impedance ratio

$$\frac{Z_0}{Z_0'} = \frac{b'}{b} - \frac{\alpha a}{b}\left(\frac{1}{2} - \frac{2}{\pi^2}\right)$$

Equating Z_0 and Z_0' we obtain

$$b' = b + a\alpha\left(\frac{1}{2} - \frac{2}{\pi^2}\right) \quad . \quad . \quad . \quad . \quad (46)$$

Thus if the hexagonal guide has dimensions a' and b', as given in eqns. (45) and (46), it has the same guide wavelength and same impedance as (and so is matched to) a rectangular guide of dimensions a and b. This enables one to design a hexagonal guide which has the same known behaviour as a rectangular one. Experimental results have been obtained which agree well with the theoretical eqns. (45) and (46).

This "similar guide" theory has been used to calculate the behaviour of a number of junctions. It is particularly applicable when we need to assess the manufacturing tolerances which are required to ensure that the behaviour of a junction lies within certain specified limits.

A feature of the definition (31) of characteristic impedance is that it is also applicable when the dominant mode is a TEM wave. A TEM wave always has a unique voltage and current, and consequently a unique characteristic impedance exists for such a wave. However, definition (31) is in agreement with the unique definition so far as relative impedances are concerned. Thus eqn. (31) leads to the correct impedance mismatch for TEM modes as well as H modes.

The most important property of definition (31) is that it so determines the relative impedances that a 2-terminal description of the junction is achieved when such a description is possible either in an exact or approximate sense. However, this property is not possessed by the usual conventional definition of the impedance of a rectangular waveguide, which defines Z_0 according to the equation

$$Z_0 = \frac{b}{a}\frac{\lambda_g}{\lambda}\sqrt{\frac{\mu}{\epsilon}} \text{ ohms} \quad . \quad . \quad . \quad . \quad (47)$$

From this we find that the ratio of the impedances of two rectangular guides of different E-plane dimensions a and a' is

$$\frac{Z_0}{Z_0'} = \frac{a'}{a}\frac{\lambda_g}{\lambda_g'}$$

This formula entirely disagrees with the result given in Section 11 [cf. eqn. (40)]. Thus if definition (47) is adopted, a 2-terminal description is not possible.

As stated previously, eqn. (31) is also applicable when the two waveguides contain different media. In fact, for two identical waveguides containing different media whose dominant modes are either H waves or TEM waves, eqn. (31) leads to the exact equations for the reflections at the interface of the media derived by conventional methods.

(14) ACKNOWLEDGMENT

The author wishes to thank Mr. S. H. Moss for suggesting the problem and for his helpful interest and guidance throughout.

(15) REFERENCES

(1) Levine, H., and Schwinger, J.: "On the Theory of Electromagnetic Diffraction by an Aperture in an Infinite Plane Conducting Screen," *The Theory of Electromagnetic Waves* (Interscience Publishers, New York, 1951).

(2) Lewin, I.: "The Advanced Theory of Waveguides" (Iliffe, 1951).

(3) Marcuvizt, N. "M.I.T. Radiation Laboratory Series, Vol. 10" (McGraw-Hill, New York, 1951), pp. 296–300.

21.372.8:621.372.43

Monograph No. 149 R
Sept. 1955

THE DESIGN OF QUARTER-WAVE MATCHING LAYERS FOR DIELECTRIC SURFACES

By R. E. COLLIN, B.Sc., Ph.D., and JOHN BROWN, M.A., Ph.D., Associate Member.

The paper was first received 6*th April, and in revised form* 25*th May*, 1955. *It was published as an* INSTITUTION MONOGRAPH, *in September*, 1955.)

SUMMARY

A quarter-wave transformer to match the junction between an empty vaveguide and one completely filled with a dielectric may be made rom a waveguide partially filled with dielectric. A method of designng such a transformer, when all the waveguides have the same crossection, is described, and experimental results are given to show that his design is satisfactory. A similar arrangement can be used to natch the surfaces of a dielectric lens: slots are cut on the surface and lesign information is given for slots parallel or perpendicular to the :lectric field of the wave incident on the surface. Measured reflection :oefficients for a surface matched in this way are in good agreement vith calculated values.

LIST OF PRINCIPAL SYMBOLS

a = Broad dimension of rectangular waveguide.
b = Narrow dimension of rectangular waveguide.
ϵ_r = Relative permittivity of dielectric in filled waveguide.
$\lambda_1, \lambda_2, \lambda_3$ = Wavelength of modes propagating in empty, partially-filled and completely filled waveguides respectively.
Z_1, Z_2, Z_3 = Wave impedances for modes propagating in empty, partially-filled and completely filled waveguides respectively.
p, q, r, s = Lengths of transmission line in equivalent circuit.
l = Length of matching step.
λ_0 = Free-space wavelength.
t = Thickness of matching step.
d_1, d_2, d_3, d_4 = Distances defined in Fig. 4.

For the slotted matching with the slots cut parallel to the direction of the electric field,

c = Slot spacing.
d = Slot depth.
e = Thickness of dielectric in slotted section.

When these three symbols are primed, they refer to slots perpendicular to the direction of the electric field.

(1) INTRODUCTION

Junctions between an empty waveguide and one of the same cross-section completely filled by a dielectric occur in the construction of dielectric rod aerials and phase changers. A matchng layer to eliminate the reflection at such a junction may be lesigned on the same principle as a transmission line quarterwave transformer. When the waveguide cross-section is rectangular, the relative permittivity of the matching layer must be such that the guide wavelength in this layer is the geometric mean of the wavelengths in the empty and filled waveguides.[1] This assumes that the dielectric for the matching layer completely fills the waveguide and that only H_{01} modes are excited. This method of matching can seldom be used in practice because suitable materials are not available for the matching layer. An alternative matching layer can be made from a waveguide partially filled with dielectric, the dimensions of the dielectric filling being calculated to give the required wavelength. A similar method may be used to cancel the reflection at a free-space/dielectric surface such as occurs in a lens aerial.

(2) EQUIVALENT CIRCUIT FOR THE MATCHING LAYER

Suppose the junction to be matched involves a rectangular waveguide of cross-section, $a \times b$, and that the dielectric in the filled region has a relative permittivity, ϵ_r. The matching layer to be considered is formed by a partially-filled waveguide with the cross-section shown in Fig. 1(a): this will be referred to, in accordance with normal waveguide nomenclature, as a symmetrical H-plane slab. This appears more complicated than the asymmetrical H-plane slab, Fig. 1(b), or the asymmetrical E-plane slab, Fig. 1(c), but neither of these is preferable. The asymmetrical H-plane slab inevitably excites the H_{02} mode in both the empty and the filled waveguides: while this mode is evanescent in the empty waveguide, it may propagate in the filled waveguide unless the relative permittivity of the dielectric is nearly unity. The symmetrical H-plane slab does not excite the H_{02} mode, and the next mode—H_{03}—is evanescent in the filled waveguide for the values of relative permittivity commonly used. The asymmetric E-plane slab is acceptable from the point of view of exciting unwanted propagating modes, but its design depends on both dimensions a and b, whereas that of the symmetrical H-plane slab depends only upon a. For these reasons, the symmetrical H-plane slab is preferred and is the only one to be considered further.

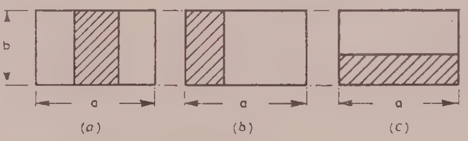


Fig. 1.—Possible arrangement of dielectric in the matching layer.

(a) Symmetrical H-plane slab.
(b) Asymmetrical H-plane slab.
c) E-plane slab.

The cross-section of the waveguide with the matching layer is shown in Fig. 2(a) and the equivalent circuit in Fig. 2(b). The calculation of the parameters of the equivalent circuit has been described in another paper[2] and detailed results have been obtained for the particular junctions of interest here.[3] The characteristic impedances of the transmission lines in the equivalent circuit are taken equal to the wave impedances of the corresponding waveguide modes. Furthermore, these wave impe-

Correspondence on Monographs is invited for consideration with a view to ublication.

Dr. Collin, who was formerly at the Imperial College, University of London, is now t C.A.R.D.E., Quebec.

Dr. Brown is in the Department of Electrical Engineering, Imperial College, Jniversity of London.

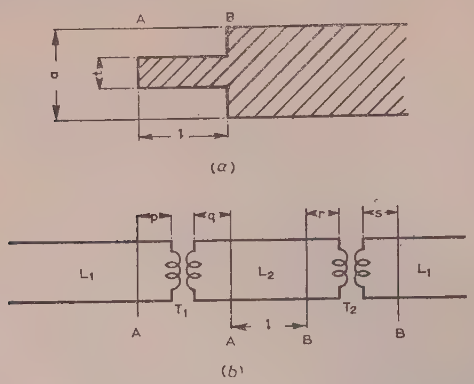


Fig. 2.—Matching layer (*a*) and equivalent circuit (*b*).

dances are directly proportional to the wavelengths of the modes,* so that

$$Z_2 = \lambda_2 Z_1/\lambda_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$Z_3 = \lambda_3 Z_1/\lambda_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The detailed calculations referred to above[3] have shown that the turns-ratios of the ideal transformers, T$_1$ and T$_2$, are within 1% of unity for relative permittivities up to 3. The extra lengths of transmission line, p, q, r and s, appear in the circuit because of reactive fields excited at the junctions, A and B [Fig. 2(*a*)]. The net phase-shift at each junction is nearly zero: e.g. at A, $[(2\pi p/\lambda_1) + (2\pi q/\lambda_2)]$ is very small. This happens because p is positive and q negative: similarly, at B, r is positive and s negative. The extra terms to be added to the length of the matching section, i.e. q and r, are of opposite sign, and since they arise from discontinuities of a similar magnitude, are likely to be approximately equal numerically. The effective electrical length of the section $(q + l + r)$ may therefore be taken to the physical length, l: once again, this assumption has been shown to be valid by a full analysis.

(3) WAVEGUIDE MATCHING LAYER

(3.1) Design Procedure

The procedure for calculating the dimensions of the symmetrical H-plane matching step is quite straightforward when the approximations discussed in Section 2 are made. The impedance of the transformer section, Z_2, must be equal to $(Z_1Z_3)^{1/2}$ so that from eqns. (1) and (2),

$$\lambda_2 = (\lambda_1\lambda_3)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The wavelengths λ_1 and λ_3 are related to the free-space wavelength, λ_0, by the equations

$$1/\lambda_1^2 = 1/\lambda_0^2 - 1/(2a)^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$1/\lambda_3^2 = \epsilon_r/\lambda_0^2 - 1/(2a)^2 \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Since the difference between the electrical and physical lengths of the matching step can be ignored,

$$l = \lambda_2/4 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The only remaining unknown is t, the thickness of the matching step, and this must be calculated to make the wavelength λ_2, have the required value, which may be done from the equation[4]

$$\left(\frac{1}{\lambda_0^2} - \frac{1}{\lambda_2^2}\right)^{1/2} \cot\left[\left(\frac{\epsilon_r}{\lambda_0^2} - \frac{1}{\lambda_2^2}\right)^{1/2} \pi t\right]$$
$$= \left(\frac{\epsilon_r}{\lambda_0^2} - \frac{1}{\lambda_2^2}\right)^{1/2} \tan\left[\left(\frac{1}{\lambda_0^2} - \frac{1}{\lambda_2^2}\right)^{1/2} (a - t)\pi\right] \quad . \quad (7)$$

The value of t is found by solving this equation numerically.

As an example, consider a waveguide for which a is 2·54 cm operating at a free-space wavelength of 3·18 cm (i.e. $a/\lambda_0 = 0\cdot8$), so that eqn. (4) gives

$$\lambda_1 = 4\cdot07$$

If the dielectric in the filled part of the waveguide has a relative permittivity of 2·47, eqns. (5) and (3) give respectively

$$\lambda_3 = 2\cdot20\text{: } \lambda_2 = 2\cdot99$$

so that by eqn. (6) $\qquad l = 0\cdot75$

Numerical solution of eqn. (7) for the above values gives

$$t = 0\cdot39$$

The design of the matching step is now completed.

The way in which the reflection coefficient of the matched junction varies with operating wavelength is shown in Fig. 3.

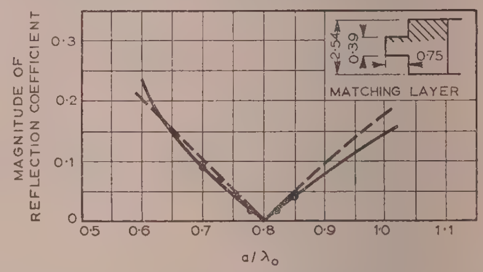


Fig. 3.—Dependence of reflection coefficient of matched waveguide junction on wavelength.

. . . Experimental points.

Both curves have been calculated for the junction specified by the above numerical values: the broken curve is based on the assumption that the wave impedances do not vary with wavelength, while the solid curve makes allowance for such variation.

(3.2) Experimental Results

A matched junction has been constructed to the dimensions calculated in Section 3.1, and measured values of the reflection coefficient are plotted in Fig. 3. The measurements have been made by a modification of the Weissfloch or tangent method.[5,6] The experimental arrangement is shown in Fig. 4(*a*), the dielectric-filled portion of the waveguide being terminated by a short-circuit. The reflection coefficient of the matched dielectric surface can be calculated from an observed curve of $[(2\pi d_1/\lambda_1) + (2\pi d_2/\lambda_3)]$ against $(2\pi d_1/\lambda_1)$. This requires the zero field position to be observed for a series of values of d_2, obtained, for example, by machining successive small pieces from the dielectric. The waveguide run must therefore be dismantled a large number of times and the measurement is a very tedious one. An equivalent result is obtained much more simply by using a short-circuiting plunger travelling in an empty waveguide, as shown in Fig. 4(*b*). The location of an effective short-circuit in

* The wave impedance of a mode propagating in a partially-filled waveguide is only proportional to the wavelength if the relative permeability within the waveguide is everywhere unity. The paper is concerned only with dielectric fillings so that this condition is satisfied. For problems involving relative permeabilities other than unity—e.g. waveguides partially filled with ferrites—the calculations must be modified to allow for the appropriate values of the wave impedances.

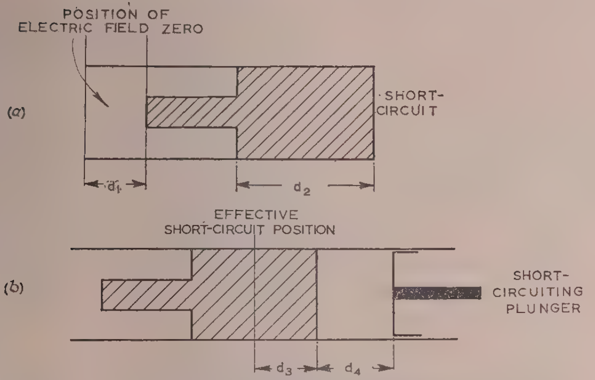
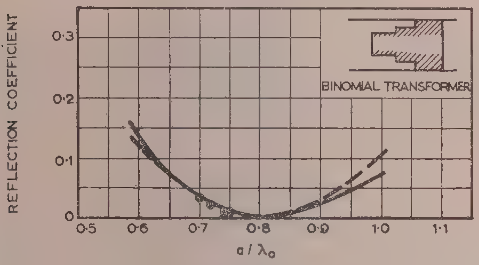


Fig. 4.—Experimental arrangement for measuring the reflection coefficient.

he dielectric-filled waveguide is deduced from the short-circuit n the empty waveguide by using the equation

$$\lambda_1 \tan (2\pi d_4/\lambda_1) = -\lambda_3 \tan (2\pi d_3/\lambda_3) \quad . \quad . \quad . \quad (8)$$

A two-step binomial transformer[7] has been designed using the ame principles and experimental and theoretical values for the eflection coefficient are compared in Fig. 5. As for the simple ransformer, the agreement is sufficiently good to establish the alidity of the design procedure. The asymmetry evident in the urves of both Figs. 3 and 5 arises because of the rapid change in he properties of the H_{01} mode in the empty waveguide as cut-off $a/\lambda_0 = 0 \cdot 5$) is approached.



ig. 5.—The reflection coefficient of a junction matched by a two-step binomial transformer.

. . . Measured values.

4) MATCHING LAYER FOR A FREE-SPACE/DIELECTRIC SURFACE

(4.1) Slots Parallel to the Direction of the Electric Field

Reflections at the surface of a dielectric lens aerial result in a irect loss of power from the radiated beam and may also cause he appearance of additional side-lobes.[8] Similar reflections in ptical lenses have been avoided by "blooming" or "coating" he surfaces, this being a direct application of the quarter-wave ransformer. The use of this technique to match the surfaces of dielectric lens is again restricted by the lack of materials with a ontinuous range of relative permittivity. An equivalent effect nay be obtained, however, by slotting the dielectric surface as n Fig. 6. The slots may be directed either parallel or perpenicular to the direction of the electric field in the wave incident n the surface, and in either case the design proceeds in a very imilar way to that given in Section 3.1. Slots parallel to the lectric-field direction are considered in this Section and those erpendicular to the electric field in Section 4.2.

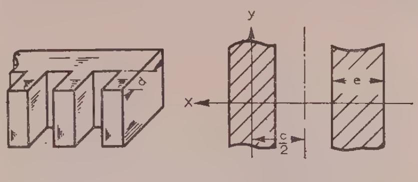


Fig. 6.—Slotted section to match dielectric/free-space boundary.

The first step in designing the slotted section is to select a suitable value for the slot spacing, c, which corresponds to the waveguide dimension a in the previous example. The only restriction on c is that no unwanted propagated waves should be excited either in free space or within the dielectric. The possible waves are the various orders of diffraction spectra for a grating of spacing c and none will propagate if c satisfies the inequality

$$c < \lambda_0/2(\epsilon_r)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

The condition given by eqn. (9) is sufficient no matter what the angle of incidence, but it may be relaxed if the range of angles of incidence is restricted. In particular, if only normally incident waves are permitted, eqn. (9) may be replaced by

$$c < \lambda_0/(\epsilon_r)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

An analysis[3] of the reflections at the junctions between free space, the slotted dielectric region and the continuous dielectric shows that the behaviour corresponds to abrupt impedance changes at each junction, and that the impedance of each section is directly proportional to the wavelength measured within it. As in the waveguide example discussed earlier, the reactive fields excited at each of the junctions may be neglected. The wavelength in the slotted section, λ_s, must satisfy the equation

$$\lambda_s = \lambda_0/(\epsilon_r)^{1/4} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

which follows from eqn. (3) since the wavelength in the continuous dielectric is $\lambda_0/(\epsilon_r)^{1/2}$. The depth of the slots, i.e. the dimension d in Fig. 6, must therefore be

$$d = \lambda_s/4 = \lambda_0/4(\epsilon_r)^{1/4} \quad . \quad . \quad . \quad . \quad . \quad (12)$$

The thickness of the dielectric in the slotted section, e, is now calculated to give the required value of λ_s and must satisfy the following equation [derived in Section (8.1)]:

$$\left(\frac{\epsilon_r}{\lambda_0^2} - \frac{1}{\lambda_s^2}\right)^{1/2} \tan\left[\left(\frac{\epsilon_r}{\lambda_0^2} - \frac{1}{\lambda_s^2}\right)^{1/2} \pi e\right] = \left(\frac{1}{\lambda_s^2} - \frac{1}{\lambda_0^2}\right)^{1/2} \tanh\left[\left(\frac{1}{\lambda_s^2} - \frac{1}{\lambda_0^2}\right)^{1/2} \pi(c-e)\right] \quad . \quad (13)$$

This equation is similar in form to eqn. (7), and when λ_s is replaced by the value from eqn. (11), it reduces to

$$(\epsilon_r)^{1/4} \tan\left\{\pi e[\epsilon_r - \sqrt{(\epsilon_r)}]^{1/2}/\lambda_0\right\} = \tanh\left\{\pi(c-e)[\sqrt{(\epsilon_r)} - 1]^{1/2}/\lambda_0\right\} \quad . \quad (14)$$

The set of curves in Fig. 7, showing e/c against c/λ_0 for various values of ϵ_r, has been computed from eqn. (14).

A matching layer has been designed to cancel the reflection when a plane wave of $\lambda = 10$ cm is normally incident upon a slab of polystyrene ($\epsilon_r = 2 \cdot 56$). The less stringent condition on c [eqn. (10)] is adequate in this case and requires that c should be less than $6 \cdot 3$ cm: 4 cm has been taken as a suitable value for c. The ratio c/λ_0 is therefore $0 \cdot 4$ and computation based on eqn. (14) gives e/c equal to $0 \cdot 34$, i.e. e is $1 \cdot 36$ cm. The depth of the slots

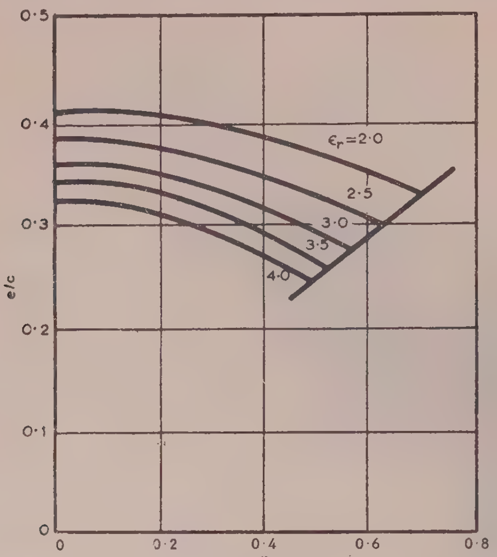


Fig. 7.—Design curves for matching layer with slots parallel to the electric field.

The straight line through the right-hand ends of the curves for ϵ_r has the equation $c/\lambda_0 = 1/\sqrt{\epsilon_r}$.

as calculated from eqn. (12) is 1·98 cm, which completes the design.

The matching layer made to this design[9] has been tested in a parallel-plate transmission line within which the field approximates closely to a plane wave. Measurements of the reflection coefficient have been carried out using virtually the same technique as described in Section 3.2. Since no movable short-circuiting plunger is available for the transmission line, the dielectric slab has to be moved relatively to a fixed short-circuit to obtain the Weissfloch curve. The calculated reflection-coefficient curve and the measured points are shown in Fig. 8.

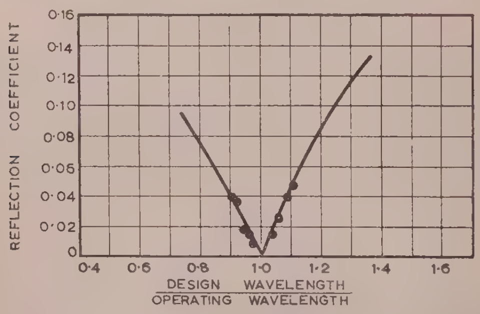


Fig. 8.—Reflection coefficient of matched free-space/dielectric surface.

. . . Measured values.

The measured values indicate that the reflection coefficient is a minimum at 9·9 cm instead of the desired 10 cm. This difference is probably due to slight departures of the field within the transmission line from an ideal plane wave. The matching layer reduces the reflected power from the dielectric surface to less than 0·25% of the incident power for wavelengths between 9 and 11·5 cm, i.e. over a band of more than ±10%. The unmatched surface reflects 5·3% of the incident power.

The radiation efficiency of a dielectric lens aerial may be considerably increased by eliminating reflections from the surfaces. In such an application, the angle of incidence at the dielectri surface may vary from zero up to about 30°, and to ensur the absence of higher-order reflected waves it is essential t make the spacing c satisfy the inequality in eqn. (19). Th reflection is completely cancelled only at normal incidence, but i considerably reduced at other angles. An approximate indica tion of the reduction may be obtained by observing that a chang from normal incidence to an angle of incidence i is equivalent t changing the wavelength from λ_0, the design value, to $\lambda_0 \sec i$ this follows from a consideration of the analogy between plan waves and waves on transmission lines. In the present example therefore, the reflected power is less than 0·25% of the inciden power provided that cos i is less than 0·9 at the design wavelength i.e. if i is less than 25°.

(4,2) Slots Perpendicular to the Direction of the Electric Field

The general considerations are exactly the same as in the pre vious Section, and in particular the restriction on the size c the spacing between the slots, c', must hold. The only differenc in the design procedure is that the wavelength in the slotte section, λ'_s, now satisfies

$$\left(\frac{\epsilon_r}{\lambda_0^2}-\frac{1}{\lambda_s'^2}\right)^{1/2}\tan\left[\left(\frac{\epsilon_r}{\lambda_0^2}-\frac{1}{\lambda_s'^2}\right)^{1/2}\pi e'\right]$$
$$=\epsilon_r\left(\frac{1}{\lambda_s'^2}-\frac{1}{\lambda_0^2}\right)^{1/2}\tanh\left[\left(\frac{1}{\lambda_s'^2}-\frac{1}{\lambda_0^2}\right)^{1/2}\pi(c'-e')\right] \quad . \quad (15$$

in place of eqn. (13). The derivation of this eqn. (15) is given i Section 8.2: the equation corresponding to eqn. (14) is

$$\tan\left\{[\epsilon_r-\sqrt{(\epsilon_r)}]^{1/2}\pi e'/\lambda_0\right\}$$
$$=(\epsilon_r)^{3/4}\tanh\left\{\pi[\sqrt{(\epsilon_r)}-1]^{1/2}[c'-e']/\lambda_0\right\} \quad . \quad (1$$

and curves of e'/c' are plotted against c'/λ_0 for various values c ϵ_r in Fig. 9.

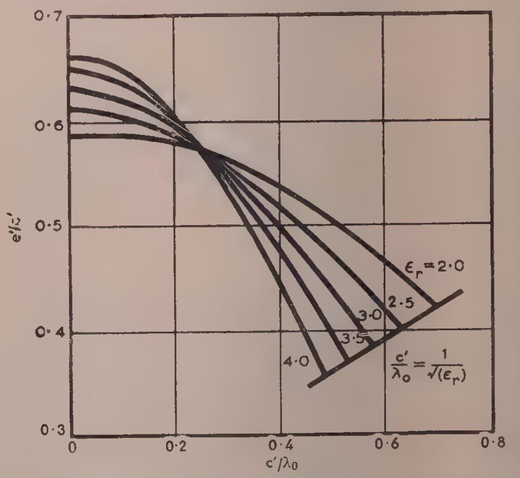


Fig. 9.—Design curves for matching layer with slots perpendicular the electric field.

In certain radar sets, circularly polarized waves are require and the aerial system must then operate for each of two pe pendicular directions of the electric field. A lens matching lay in such a case must be effective for both directions of the electr field, and an obvious way of achieving this is to cut two pe pendicular sets of slots of equal spacings and slot widths. It is doubtful whether the wavelength in such a structure could be computed, so that the design would have to be found by experiment. The possibility that one set of slots can be used as a matching layer for plane waves with the electric field either parallel or perpendicular to the direction of the slots has therefore been investigated. There is no design which matches both polarizations simultaneously, but by choosing a slot width which is the average of those required for the two different polarizations a reasonable performance may be obtained. This is illustrated in Fig. 10, which shows the calculated variation of reflection coefficient with wavelength for a matching layer designed in this way: the design wavelength is 10 cm, the slot spacing is 4 cm, and the thickness of the dielectric is 1·75 cm, i.e. the average of 1·36 cm and 2·14 cm, which are the values required for plane waves with the electric field parallel and perpendicular to the slot respectively.

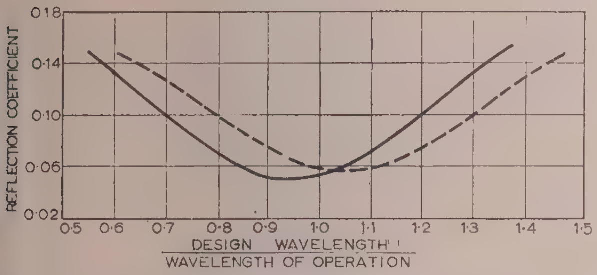


Fig. 10.—Reflection coefficient of matched free-space/dielectric surface.
—— Electric field parallel to slots.
---- Magnetic field parallel to slots.

(5) CONCLUSIONS

Matching sections for dielectric surfaces may be made by cutting slots of suitable size: the dimensions of these slots can be calculated from the wavelength in the slotted section and formulae are given for slots parallel or perpendicular to the incident electric field. Measurements have confirmed the validity of the design procedure.

(6) ACKNOWLEDGMENT

Dr. Collin is indebted to the Athlone Committee for a Fellowship which was held while the work described in the paper was carried out.

(7) REFERENCES

(1) RAMO, S., and WHINNERY, J. R.: "Fields and Waves in Modern Radio" (Wiley and Sons, New York).
(2) COLLIN, R. E., and BROWN, J.: "The Calculation of the Equivalent Circuit of an Axially Unsymmetrical Waveguide Junction" (see page 121).
(3) COLLIN, R. E.: "Interface Problems at Dielectric Discontinuities in Waveguides." Thesis submitted for Doctor of Philosophy Degree, University of London.
(4) MONTGOMERY, C. G., DICKE, R. H., and PURCELL, E. M.: "Principles of Microwave Circuits," Radiation Laboratory Series: No. 8 (McGraw-Hill, 1948), p. 388.
(5) MARCUVITZ, N.: "Waveguide Handbook" (McGraw-Hill, 1951), Section 3.4.
(6) BARLOW, H. E. M., and CULLEN, A. L.: "Microwave Measurements" (Constable, 1950), Section 6.5.
(7) SOUTHWORTH, C. G.: "Principles and Application of Waveguide Transmission" (D. Van Nostrand, 1950).
(8) BROWN, J.: "Microwave Lenses" (Methuen, 1953).
(9) EL-KHARADLY, M. M. Z.: "Some Experiments on Artificial Dielectrics at Centimetric Wavelengths," *Proceedings I.E.E.*, Paper No. 1700 R, January, 1955 (**102 B**, p. 17).

(8) APPENDICES: THE WAVELENGTH WITHIN THE SLOTTED REGION

(8.1) Electric Field Parallel to the Slots

The system of co-ordinates used is shown in Fig. 6, the z-axis being into the plane of the paper. When the electric field is parallel to the slots, the only field components which are excited are E_y, H_x and H_z. The two magnetic-field components are related to E_y by

$$j\omega\mu_0 H_x = \partial E_y/\partial z : j\omega\mu_0 H_z = -\partial E_y/\partial x \quad . \quad . \quad (16)$$

both in free space and within the dielectric.

The electric field must satisfy the differential equations

$$\frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial z^2} + k_0^2 E_y = 0:\ \frac{e}{2} \leqslant |x| \leqslant \frac{c}{2} \quad . \quad . \quad (17)$$

$$\frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial z^2} + \epsilon_r k_0^2 E_y = 0:\ 0 \leqslant |x| \leqslant \frac{e}{2} \quad . \quad . \quad (18)$$

where $k_0 = 2\pi/\lambda_0$, the free-space propagation coefficient.

The fields are periodic with the spacing c and are symmetrical about the plane $x = 0$, so that

$$\partial E_y/\partial x = 0 \text{ for } x = 0 \text{ and } x = c/2 \quad . \quad . \quad . \quad (19)$$

Furthermore, E_y and H_z, i.e. $\partial E_y/\partial x$, must be continuous over the boundary plane $x = e/2$.

If $\beta = 2\pi/\lambda_s$ is the phase coefficient in the direction of propagation, the z-axis, then from eqns. (17)–(19)

$$\left.\begin{aligned} E_y &= A \cos\left[(\epsilon_r k_0^2 - \beta^2)^{1/2} x\right] \exp(-j\beta z):\ 0 \leqslant x \leqslant e/2 \\ &= B \cosh\left[(\beta^2 - k_0^2)^{1/2}\left(\frac{c}{2} - x\right)\right] \exp(-j\beta z):\\ &\qquad\qquad e/2 \leqslant x \leqslant \frac{c}{2} \end{aligned}\right\} (20)$$

From the continuity of E_y at $x = e/2$

$$A \cos\left[(\epsilon_r k_0^2 - \beta^2)^{1/2} e/2\right] = B \cosh\left[(\beta^2 - k_0^2)^{1/2}\left(\frac{c}{2} - \frac{e}{2}\right)\right] \ . \ (21)$$

and from the continuity of $\partial E_y/\partial x$

$$(\epsilon_r k_0^2 - \beta^2)^{1/2} A \sin\left[(\epsilon_r k_0^2 - \beta^2)^{1/2} e/2\right]$$
$$= (\beta^2 - k_0^2)^{1/2} B \sinh\left[(\beta^2 - k_0^2)^{1/2}\left(\frac{c}{2} - \frac{e}{2}\right)\right] \quad . \quad (22)$$

Eliminate A and B from eqns. (21) and (22).

Then $(\epsilon_r k_0^2 - \beta^2)^{1/2} \tan\left[(\epsilon_r k_0^2 - \beta^2)^{1/2} e/2\right]$

$$= (\beta^2 - k_0^2)^{1/2} \tanh\left[(\beta^2 - k_0^2)^{1/2}\left(\frac{c}{2} - \frac{e}{2}\right)\right] \ . \ (23)$$

which becomes eqn. (13) on substitution for k_0 and β.

(8.2) Electric Field Perpendicular to the Slots*

The field components are now H_y, E_x and E_z and

$$j\omega\epsilon_0\epsilon_r E_x = -\partial H_y/\partial z:\ j\omega\epsilon_0\epsilon_r E_z = \partial H_y/\partial x \text{ if } |x| \leqslant e'/2 \ . \ (24)$$

* Primes are attached to those quantities which have values different from the ones in the preceding Section.

$$j\omega\epsilon_0 E_x = -\partial H_y/\partial z: \; j\omega\epsilon_0 E_z = \partial H_y/\partial x \text{ if } e'/2 \leqslant |x| \leqslant c'/2 \quad (25)$$

H_y satisfies wave equations and boundary conditions similar in form to eqns. (17)–(19) so that

$$H_y = A' \cos\,[(\epsilon_r k_0^2 - \beta'^2)^{1/2} x]: \; |x| \leqslant e'/2 \quad (26)$$

$$= B' \cosh\,[(\beta'^2 - k_0^2)^{1/2}(c'/2 - x)]: \; e'/2 \leqslant |x| \leqslant c'/2 \quad (27)$$

On the boundary $x = e'/2$, H_y and E_z are continuous so that

$$A' \cos\,[(\epsilon_r k_0^2 - \beta'^2)^{1/2} e'/2] = B' \cosh\left[(\beta'^2 - k_0^2)^{1/2}\left(\frac{c'}{2} - \frac{e'}{2}\right)\right] \quad (28)$$

and

$$\frac{A'}{\epsilon_r}(\epsilon_r k_0^2 - \beta'^2)^{1/2} \sin\,[(\epsilon_r k_0^2 - \beta'^2)^{1/2} e'/2] = B'(\beta'^2 - k_0^2)^{1/2} \sinh\left[(\beta'^2 - k_0^2)^{1/2}\left(\frac{c'}{2} - \frac{e'}{2}\right)\right] \quad (29)$$

Eliminate A' and B', so that

$$(\epsilon_r k_0^2 - \beta'^2)^{1/2} \tan\,[(\epsilon_r k_0^2 - \beta'^2)^{1/2} e'/2] = \epsilon_r(\beta'^2 - k_0^2)^{1/2} \tanh\left[(\beta'^2 - k_0^2)^{1/2}\left(\frac{c'}{2} - \frac{e'}{2}\right)\right] \quad (30)$$

Substitution for β' and k_0 gives eqn. (15).

621.3.011: 538.3

Monograph No. 150
Sept. 1955

ELECTROMAGNETIC MOMENTUM AND ELECTRON INERTIA IN A CURRENT CIRCUIT

By Professor E. G. CULLWICK, O.B.E., M.A., D.Sc., F.R.S.E., Member.

(*The paper was first received* 12*th April, and in revised form* 13*th June*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in September*, 1955.)

SUMMARY

In the second volume of his "Treatise on Electricity and Magnetism" Clerk Maxwell developed the theory of electric current-circuits from general dynamical principles, and discussed the experimental effects which should occur if an electric current is a true motion of some substance possessing inertia. Since none of these effects had at that time been observed, Maxwell developed his general electromagnetic theory on the assumption that they do not exist, or at least that they produce no sensible effect.

It is now known, however, that an electric current in a conductor consists of moving electrons, and the inertia effects which were discussed by Maxwell have been observed experimentally. They are extremely small, and have not been brought within the scope of electromagnetic theory. A conduction current is usually assumed to be due to the drifting along the conductor, with a very small mean velocity, of all the available conduction electrons, so that the kinetic energy of the electrons due to this motion is negligible in comparison with the magnetic energy of the current. Electron-inertia effects in current circuits have therefore been accepted as something outside classical electromagnetic theory—a position which is illogical if, as is usual, we identify the kinetic and magnetic energies of a free electron.

It is shown in the paper that it is possible to identify the kinetic energy of the conduction electrons in a current circuit with the magnetic energy of the current, so that electron-inertia effects can be included in the general electromagnetic scheme. In consequence, a current circuit can be said to possess an electromagnetic mass whose motion, when current flows, entails electromagnetic momentum. This momentum accounts for the known effects of electron inertia and also for the force on the end wire of a long rectangular circuit.

The relativistic form of the theory indicates the possibility that electromagnetic laws may depart from the classical form, becoming non-linear in circuits where a high inductance per unit length of conductor is combined with a current greater than is usually found in practice.

The inadequacy of classical theory also extends to the known electromagnetic properties of superconductors, and the present hypothesis suggests the possibility of a unified theory in which there would be no necessity to distinguish between a superconductor and a perfect conductor.

LIST OF SYMBOLS

(Rationalized M.K.S. units)

$a = 1 \cdot 7 \times 10^{-3} l/L$.
$A, \mathbf{A}$ = Vector potential, webers/m.
$B, \mathbf{B}$ = Magnetic flux density, webers/m^2.
c = Velocity of light *in vacuo*, m/sec.
$D, \mathbf{D}$ = Electric flux density, coulombs/m^2.
e = Electronic charge (a negative quantity), coulombs.
$E, \mathbf{E}$ = Electric field intensity, volts/m.
V_e = E.M.F., volts.
F = Mechanical force, newtons.
$H, \mathbf{H}$ = Magnetic field intensity, AT/m.
I, i = Electric current, amp.
$J, \mathbf{J}$ = Electric current density, amp/m^2.
L = Total self-inductance of a current circuit, henrys.
L_0 = Self-inductance per unit length of a 2-wire transmission line, henrys/m.
l = Length of circuit, m.
$m, \ M$ = Mass, kg.
m_0 = Rest mass of an electron, kg.
M_0, M = Electromagnetic mass of a current circuit, kg.
N = Numbers of turns; effective number of conduction electrons per unit length of a conductor.
$p, \mathbf{p}$ = Electromagnetic momentum per unit volume of the field.
p_{total} = Total electromagnetic momentum.
q, Q = Electric charge, coulombs.
$R, \mathbf{r}$ = Radius, radius vector, m.
r = Resistivity, ohm-m.
R = Resistance, ohms.
$S, \mathbf{S}$ = Poynting vector.
t = Time, sec.
T_{me} = Mutual kinetic energy, current and conductor.
U, W = Energy, joules.
u, v, w = Velocity, m/s.
x, z = Co-ordinates.
α, θ = Angles.
$\beta = \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}}$
ϵ_0 = Primary electric constant,* $8 \cdot 854 \times 10^{-12}$.
μ_0 = Primary magnetic constant,* $1 \cdot 257 \times 10^{-6}$.
ρ = Charge of effective conduction electrons per unit length of conductor (a negative quantity), coulombs/m.
Φ = Magnetic flux linkage, weber-turns.
ψ = Scalar potential function.

(1) INTRODUCTION

It is a common practice in electromagnetic theory to regard the magnetic energy of a current circuit as electrokinetic, and to compare the expression $\frac{1}{2}LI^2$ with the kinetic energy of a moving mass, $\frac{1}{2}mv^2$. It is the purpose of the paper to show that the magnetic energy of a current circuit can be identified with the kinetic energy of the mass-equivalent of the total electromagnetic energy of the conduction electrons. The concept of electromagnetic momentum in a current circuit will then be used to determine the force on the end wire of a long rectangular circuit, and to bring the known effects of electron inertia in a circuit within the scope of electromagnetic theory.

(2) ELECTROMAGNETIC MASS OF A MOVING CHARGED PARTICLE

Fig. 1 shows a small positive charge, q, moving in free space with a constant velocity v in a straight line. The electromagnetic field at a point P(r, α) consists of a radial electric field of intensity E and flux density $D = \epsilon_0 E$, together with a magnetic field whose

Correspondence on Monographs is invited for consideration with a view to publication.

Professor Cullwick is Watson-Watt Professor of Electrical Engineering, St. Andrews University (Queen's College, Dundee).

* The idea that "free space" possesses the physical properties of permittivity and permeability is based on the concept of a material medium or aether, and is clearly incompatible with the hypothesis of this paper.

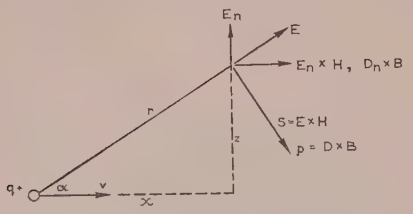


Fig. 1.—Moving charge: electromagnetic momentum and energy flux.

lines of force are circles concentric with the path of the charge. At P the magnetic field is vertically out of the paper, with intensity $\boldsymbol{H}$ and flux density $\boldsymbol{B} = \mu_0 \boldsymbol{H}$.

If the charged particle is spherical when at rest, and the velocity v is very small compared with c, we have

$$\boldsymbol{E} = \frac{q}{4\pi\epsilon_0 r^3}\boldsymbol{r} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and

$$\boldsymbol{H} = \boldsymbol{v} \times \boldsymbol{D}$$

or

$$H = v\epsilon_0 E_n = \left(\frac{q \sin\alpha}{4\pi r^2}\right)v \quad . \quad . \quad . \quad . \quad (2)$$

where E_n is the component of $\boldsymbol{E}$ perpendicular to the velocity. The flux density is

$$\boldsymbol{B} = \mu_0 \boldsymbol{H} = \frac{\boldsymbol{v} \times \boldsymbol{E}}{c^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

If v is comparable with c the particle suffers the Lorentz contraction, and becomes an oblate spheroid whose polar axis is $(1 - v^2/c^2)^{1/2}$ times its equatorial axis. The electric field at P is still radial, but has the value

$$\boldsymbol{E} = \frac{q}{4\pi\epsilon_0 r^3} \frac{1 - \dfrac{v^2}{c^2}}{\left(1 - \dfrac{v^2}{c^2}\sin^2\alpha\right)^{3/2}} \boldsymbol{r} \quad . \quad . \quad . \quad (4)$$

This has components:

normal to v:
$$E_n = \frac{q}{4\pi\epsilon_0} \frac{\beta z}{(\beta^2 x^2 + z^2)^{3/2}} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

parallel to v:
$$E_v = \frac{q}{4\pi\epsilon_0} \frac{\beta x}{(\beta^2 x^2 + z^2)^{3/2}} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where $\beta = (1 - v^2/c^2)^{-1/2}$. The magnetic field is still given by $\boldsymbol{H} = \boldsymbol{v} \times \boldsymbol{D}$, $\boldsymbol{B} = \boldsymbol{v} \times \boldsymbol{E}/c^2$, or

$$H = v\epsilon_0 E_n, \quad B = \frac{vE_n}{c^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

According to classical electromagnetic theory we also have, at the point P,

(*a*) a flux of electromagnetic energy given by the Poynting vector

$$\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

which is the rate at which energy passes through unit area normal to $\boldsymbol{S}$, and

(*b*) an electromagnetic momentum

$$\boldsymbol{p} = \boldsymbol{D} \times \boldsymbol{B} = \frac{\boldsymbol{E} \times \boldsymbol{H}}{c^2} = \frac{\boldsymbol{S}}{c^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

per unit volume of the field.

Both $\boldsymbol{S}$ and $\boldsymbol{p}$ are shown in Fig. 1. They are directed inwards towards the path of the charge, at an angle $\pi/2 - \alpha$ to the direction of motion.

The classical interpretation of eqn. (8) is that, as the position of the charged particle changes, the electromagnetic energy at stationary points in the medium or aether is maintained at appropriate values by the flow, with velocity c, of energy in the medium. There is, however, an alternative viewpoint which merits investigation. Since the velocity of the charge is constant there is no change in the total energy of the system, and for an observer moving with the particle there is no flow of electromagnetic energy at all. If, therefore, the electromagnetic energy is regarded as belonging to the particle, rather than to the medium, it is reasonable to suppose that when the particle moves it takes the energy with it, with its own velocity v.

If we define the *electromagnetic mass* of the particle, M_e, by the relation (for $v \ll c$)

$$\tfrac{1}{2}M_e v^2 = \text{magnetic energy} \quad . \quad . \quad . \quad (10)$$

it is easy to show that

$$M_e = \frac{\mu_0 q^2}{6\pi R} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where R is the radius of the charged sphere. Now let m_e be the electromagnetic mass per unit volume of the field, and suppose this mass to move, with the charge, with velocity v. We then have

$$\left.\begin{aligned} \tfrac{1}{2}m_e v^2 &= \tfrac{1}{2}HB \\ \text{so that} \quad m_e &= \frac{HB}{v^2} = \frac{\mu_0 H^2}{v^2} = \frac{E_n D_n}{c^2} \end{aligned}\right\} \quad . \quad . \quad . \quad (12)$$

the last expression arising from eqn. (7). The momentum of m_e is clearly

$$m_e v = \frac{vE_n D_n}{c^2} = D_n B \quad . \quad . \quad . \quad . \quad . \quad (13)$$

or

$$\boldsymbol{p} = \boldsymbol{D}_n \times \boldsymbol{B} \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

which is the component of the classical electromagnetic momentum, $\boldsymbol{D} \times \boldsymbol{B}$, in the direction of the motion of the charged particle.

If we similarly resolve the Poynting vector $\boldsymbol{S}$, we find the flux of energy in the direction of motion to be

$$S_v = E_n H = \frac{H^2}{\epsilon_0 v}, \quad \text{or } \boldsymbol{S}_v = \boldsymbol{E}_n \times \boldsymbol{H} \quad . \quad . \quad (15)$$

We next suppose that this represents the motion of electromagnetic energy with velocity v, moving with the charge. Its volume density will be

$$U = \frac{S_v}{v} = \frac{H^2}{\epsilon_0 v^2} = E_n D_n \quad . \quad . \quad . \quad . \quad (16)$$

This is obviously not the magnetic energy, $\tfrac{1}{2}HB$, for it is far greater. But if we use the mass-energy equivalence we find a direct relationship between these two energy densities. The mass of the energy U is

$$\frac{U}{c^2} = \frac{E_n D_n}{c^2} = m_e \quad . \quad . \quad . \quad . \quad . \quad (17)$$

the electromagnetic mass as defined by eqn. (12).

We therefore see that

(*a*) The magnetic energy of the moving charged particle can be completely identified with the kinetic energy of the mass of the moving energy U.

(*b*) The electromagnetic momentum in the direction of motion is merely the momentum of the mass of the energy U.

It should, however, be noted that eqns. (12)–(17) cannot be regarded as providing an accurate microscopic account of the energy of the field. The expression for m_e given by eqn. (12)

loes not possess spherical symmetry about the particle, but ymmetry about the direction of motion, so that it cannot be in accurate detailed description of the energy-mass distribution of a stationary spherical charge. The quantities m_e, p, S_v and U nust therefore be regarded as functions which possess physical dentity only when integrated throughout the complete volume of the field.

Since we assume that v is small compared with c, we neglect n the first instance the mass of the kinetic energy. This involves relativistic modification, which will be considered later. If he electromagnetic mass M_e is the only mass possessed by the particle (e.g. if the particle is an electron and we assume that its nass is entirely electromagnetic), then clearly the energy U is he total mass-energy of the particle, or the energy which would oe liberated if the particle disintegrated into electromagnetic adiation.

If a charged sphere is stationary or moving with a velocity $\ll c$, and if the charge is uniformly distributed on the sphere, he energy of its electric field is easily found to be

$$W_s = \frac{q^2}{8\pi\epsilon_0 R} \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

vhich has a mass-equivalent

$$M_s = \frac{W_s}{c^2} = \frac{\mu_0 q^2}{8\pi R} \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Thus $M_s = 3M_e/4$, so that if an electron is to be regarded as a pherical aggregate of charge whose mass is entirely electronagnetic, it is necessary to postulate additional electromagnetic nergy of amount

$$\frac{q^2}{24\pi\epsilon_0 R} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

t is well known that a stable spherical electron cannot exist vithout internal forces in addition to those of classical theory, nd eqn. (20) may be taken to represent the energy of the nonlassical, and at present unknown, forces which keep the electron rom exploding.

If we identify the mass, M_e, given by eqn. (11) with the rest nass m_0 of an electron, and then add the mass, m_m, of the nagnetic energy $\frac{1}{2}M_e v^2$, we obtain the total mass of the moving article

$$\begin{aligned} m = m_0 + m_m &= M_e + M_e\left(\frac{v^2}{2c^2}\right) \\ &= m_0\left(1 + \frac{v^2}{2c^2}\right) \quad . \quad . \quad . \quad (21) \end{aligned}$$

This is a first approximation, with $v \ll c$, to the relativistic elation

$$m = m_0(1 - v^2/c^2)^{-1/2} \quad . \quad . \quad . \quad . \quad . \quad (22)$$

(3) TWO MOVING CHARGED PARTICLES

(3.1) Energy Fluxes

Fig. 2 shows two particles with charges q_1 and q_2 moving vith velocities v_1 and v_2, respectively, in paths which are not ecessarily co-planar. At a point P the electric field is made up f components E_1 from q_1 and E_2 from q_2, and the magnetic eld similarly has components $H_1 = v_1 \times D_1$, $H_2 = v_2 \times D_2$. he resultant magnetic intensity is

$$H = H_1 + H_2, \quad \text{where } H^2 = H_1^2 + H_2^2 + 2H_1 \,.\, H_2 \quad (23)$$

he magnetic energy in the field has a volume density

$$W_m = \frac{\mu_0 H^2}{2} = \frac{\mu_0}{2}(H_1^2 + H_2^2 + 2H_1 \,.\, H_2) \quad . \quad (24)$$

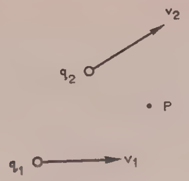


Fig. 2.—Two moving charges.

and since the mutual energy $\mu_0 H_1 \,.\, H_2$ is symmetrical in H_1 and H_2, we shall assume that equal portions of it are associated with each charge.

We therefore suppose that each charge carries its share of the mutual energy, so that using the relation given in eqn. (15) we attribute to q_1, in addition to the flux of its self-energy, a flux of energy with velocity v_1,

$$S_{m1} = \frac{H_1 \,.\, H_2}{\epsilon_0 v_1} = \frac{(v_1 \times D_1) \,.\, H_2}{\epsilon_0 v_1} = \frac{(E_1 \times H_2) \,.\, v_1}{v_1} \quad (25)$$

and to q_2 a flux of energy with velocity v_2

$$S_{m2} = \frac{H_1 \,.\, H_2}{\epsilon_0 v_2} = \frac{(v_2 \times D_2) \,.\, H_1}{\epsilon_0 v_2} = \frac{(E_2 \times H_1) \,.\, v_2}{v_2} \quad (26)$$

If v_1 and v_2 are parallel or anti-parallel, these reduce to the simple forms

$$S_{m1} = E_{n1} \times H_2, \quad S_{m2} = E_{n2} \times H_1 \quad . \quad . \quad . \quad (27)$$

where E_{n1} and E_{n2} are the components of E_1 and E_2 perpendicular to the velocities.

In addition, the charge q_1 carries a self-energy flux $S_{s1} = E_{n1} \times H_1$ with velocity v_1, and q_2 carries a self-energy flux $S_{s2} = E_{n2} \times H_2$ with velocity v_2. So the total energy flux carried by each charge is

$$\left.\begin{aligned} S_1 = S_{s1} + S_{m1} = E_{n1} \times H \\ S_2 = S_{s2} + S_{m2} = E_{n2} \times H \end{aligned}\right\} \quad . \quad . \quad . \quad (28)$$

where $H = H_1 + H_2$, the resultant magnetic intensity. These energy fluxes are assumed to exist entirely independently of each other. Even if v_1 and v_2 have different values, or are in different directions, the two energy streams are assumed not to interfere.

If the two charges are moving with the same velocity, their individual energy fluxes can be combined into one, of value

$$S = (E_{n1} + E_{n2}) \times H = E_n \times H \quad . \quad . \quad . \quad (29)$$

and since this combination can be continued indefinitely for additional charges moving with the same velocity, it follows that eqn. (29) applies to any rigid configuration of charges in uniform rectilinear motion, and that eqn. (28) is valid for the two lines of moving charge in two parallel conductors.

(3.2) The Effective Mass of a Conduction Electron

Experiments on electron inertia in closed conducting circuits (see Section 11.2) have shown that the ratio e/m for a conduction electron in a current-carrying conductor is approximately the same as that obtained in experiments on low-energy cathode rays, i.e. the effective mass of a conduction electron is approximately the same as the rest mass of a free electron.

According to Section 3.1, however, the mass of the particle q_1 is a function of the mutual energy of q_1 and q_2 as well as of its self-energy. Let us examine the probable order of magnitude of this effect for a conduction electron.

First suppose that the component of magnetic field, H_2, due to sources other than q_1 is uniform. Then it is clear that the total volume integral of $H_1 . H_2$ is zero. Consider two points diametrically opposite on a circular line of force of H_1: the values of $H_1 . H_2$ at these two points are equal and opposite and their contributions to the volume integral cancel. Thus the electromagnetic mass of an electron moving in a uniform external magnetic field is the same as that of an isolated electron.

Next consider the orders of magnitude of H_1^2 and $H_1 . H_2$ at points in the vicinity of a spherical electron. The portion of the electromagnetic mass of a charged sphere contained within a concentric sphere of radius r is

$$m_r = \frac{\mu_0 q^2}{6\pi}\left(\frac{1}{R} - \frac{1}{r}\right)$$

Let $r = 1\,000R$. Then 99·9% of the electromagnetic mass is contained in the concentric sphere. The magnetic flux-density at a radius r on the equatorial plane of the moving electron, taking $R = 2 \times 10^{-15}$ m and $r = 2 \times 10^{-12}$ m, is

$$B_1 = \frac{\mu_0 ev}{4\pi r^2} = (4 \times 10^{-3})v \text{ weber/metre}^2$$

Suppose a second electron to be moving with the same velocity in a parallel path, with a common equatorial plane, at a distance of 10^{-10} m (the order of magnitude of the diameter of an atom) from q_1. Then its magnetic field at the same point as before is about $B_2 = (1{\cdot}6 \times 10^{-6})v$ weber/m², so that $B_1^2/B_1B_2 = H_1^2/H_1H_2 \simeq 2500$.

Nearer to q_1 this ratio will be greater. It is therefore evident that the effective electromagnetic mass of a conduction electron will not differ, according to this hypothesis, appreciably from that of a free electron.

(4) APPLICATION TO A LONG RECTANGULAR CIRCUIT

(4.1) General Considerations

Fig. 3 shows a long rectangular circuit, ABCD, which carries a constant current I. The sides AB, BC and CD are connected rigidly together, but the end wire DA makes contact with the rest of the circuit by means of mercury cups, so that the force on this member can be measured. Then the force on this end wire can be calculated from a knowledge of the magnetic field, at points on DA, due solely to the current in the three sides AB, BC and CD, and calculated from the law of Biot and Savart for the magnetic field of a current element:

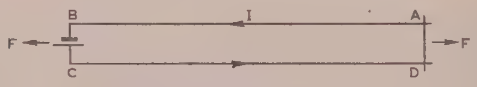


Fig. 3.—Long rectangular circuit.

$$\delta H = \frac{I}{4\pi r^3}(\delta s \times r) \quad . \quad . \quad . \quad . \quad (30)$$

where $I\delta s$ is the current element and r is the radius vector from the element to the point where the field is δH. The force on a current element $I'\delta s'$ at the point is then $F = \mu_0 I'(\delta s' \times \delta H)$, which is not, in general, equal to the force on $I\delta s$ due to the magnetic field of the element $I'\delta s'$. So the resultant force on the three-sided portion AB, BC, CD is not that which would be calculated from the magnetic field of the other portion, DA, alone, acting on the current in the three sides. Since the force on each side is perpendicular to the wire, the resultant force on the three sides is merely that on BC. But the magnetic field at BC due to DA alone, as calculated from eqn. (30), is very small and approaches zero as the length of the circuit increases. Thus, if the circuit is very long the force on the end wire DA is due entirely to the magnetic field of the two long sides, but the forces on these long sides are not due to the magnetic field of the current in DA alone.

If, as in Maxwell's original theory, the electromagnetic field is regarded as a physical condition in a material medium or aether, the reaction of the force on the end wire is considered to be borne by the medium in the vicinity and transmitted by stresses in the medium to the other end wire. It is, in fact, only since the concept of a material medium has fallen into obsolescence that problems of this kind have caused discussion,[1-7] for there appears to be nothing to take the place of the medium as an agent for transmitting the force. We shall show that such an agent is provided by the momentum of moving energy.

First let us calculate the force on the end wire by the usual method. Assume the length of the circuit to be great in comparison with its width, and let the self-inductance per unit length (two wires) at points remote from the ends be L_0. Suppose the wire DA to move a small distance δx under the action of the force F, thus increasing the length of the circuit by δx. Assume that the current I is maintained constant during the displacement. Then the magnetic flux linking the current increases by $\delta\Phi = L_0 I\delta x$, so that an e.m.f. $V_e = -\delta\Phi/\delta t$ is induced, and in order to keep the current constant additional energy must be supplied, from the source, of amount

$$\delta W = (-V_e)I\delta t = L_0 I^2 \delta x$$

The magnetic energy of the circuit increases by an amount $\frac{1}{2}L_0 I^2 \delta x$, and since the energy supplied must be equal to the mechanical work done by the force F plus the increment in magnetic energy, we have

$$\left.\begin{aligned} L_0 I^2 \delta x &= F\delta x + \tfrac{1}{2}L_0 I^2 \delta x \\ F &= \tfrac{1}{2}L_0 I^2 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (31)$$

or, if L is the total inductance of the circuit

$$F = \tfrac{1}{2}I^2\frac{dL}{dx} \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

(4.2) The Flow of Electromagnetic Momentum

In order to interpret the force F given by eqn. (32) in terms of energy momentum, we distinguish the components of electric field arising from the positive and negative charges in the conductors. We shall simplify the problem by assuming that the parallel wires have the same uniform section and negligible resistance, so that the electric field between them, except near the ends, is perpendicular to the current flow and the potential difference between them is constant. The end wire then provides a resistance "load" on the long transmission line.

Fig. 4(*a*) shows a portion of the circuit remote from the ends, and Fig. 4(*b*) is a cross-sectional view. We consider each conductor to contain the same quantity of stationary positive charge in its structure per unit length. The current in conductor 1 is provided by the motion, with a mean velocity v_1, of conduction electrons whose charge per unit length is ρ_1, and the current in conductor 2 is provided by the motion, with mean velocity v_2, of conduction electrons whose charge per unit length is ρ_2. The direction of motion is, of course, opposite to that of the current, and ρ_1 and ρ_2 are very nearly equal. Clearly $\rho_1 v_1 = \rho_2 v_2 = I$.

Then at any point P we recognize four components of electric field, all perpendicular to the wires, namely

E_1 from the conduction electrons in conductor 1.
E_2 from the conduction electrons in conductor 2.
E_3 from the stationary positive charge in conductor 1.
E_4 from the stationary positive charge in conductor 2.

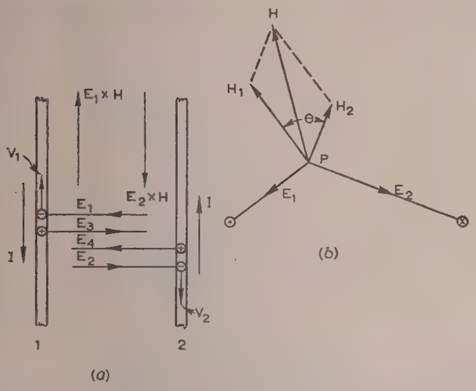


Fig. 4.—Energy fluxes in long rectangular circuit.

(a) Energy streams and electric field components.
(b) Electric and magnetic field components.

These components are in the direction shown in Fig. 4(a). The components of the magnetic field intensity, arising from the motion of the sources of E_1 and E_2, are shown in Fig. 4(b).

The positive charges, being stationary, carry no energy fluxes. The conduction electrons in conductor 1 carry an energy flux with velocity v_1:

$$S_1 = E_1 \times (H_1 + H_2)$$

or $$S_1 = |E_1 \times H_1| + |E_1 \times H_2| = E_1(H_1 + \mathrm{H}_2 \cos\theta) \quad (33)$$

directed towards the end DA.

But $E_1 = \dfrac{H_1}{\epsilon_0 v_1}$, so that

$$S_1 = \frac{(H_1^2 + H_1H_2\cos\theta)}{\epsilon_0 v_1} \quad . \quad . \quad . \quad . \quad (34)$$

Since the velocity of this energy flux is v_1, its volume density is S_1/v_1 and its mass per unit volume is

$$m_1 = \frac{S_1}{v_1c^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

Its kinetic energy is therefore

$$\tfrac{1}{2}m_1v_1^2 = \frac{\mu_0}{2}(H_1^2 + H_1H_2\cos\theta) \quad . \quad . \quad . \quad (36)$$

Similarly, the conduction electrons in conductor 2 carry an energy flux with velocity v_2:

$$S_2 = \frac{H_2^2 + H_1H_2\cos\theta}{\epsilon_0 v_2} \quad . \quad . \quad . \quad . \quad (37)$$

away from the end DA, and its kinetic energy is

$$\frac{\mu_0}{2}(H_2^2 + H_1H_2\cos\theta) \quad . \quad . \quad . \quad . \quad (38)$$

Thus the total kinetic energy of the moving energy fluxes, per unit volume, is

$$\tfrac{1}{2}\mu_0H^2 = \tfrac{1}{2}HB \quad . \quad . \quad . \quad . \quad . \quad (39)$$

i.e. the magnetic energy density.

Consider the change of electromagnetic momentum which occurs at the end wire DA. The momentum of the S_1 stream, per unit volume, is

$$S_1/c^2 = \mu_0(H_1^2 + H_1H_2\cos\theta)/v_1 \quad . \quad . \quad . \quad (40)$$

and this is deflected at the end wire at the rate

$$\mu_0(H_1^2 + H_1H_2\cos\theta) \quad . \quad . \quad . \quad . \quad (41)$$

per unit area of an infinite plane, remote from the ends, perpendicular to the side wires. So the total loss of momentum per second is

$$\mu_0\iint(H_1^2 + H_1H_2\cos\theta)ds \quad . \quad . \quad . \quad (42)$$

the integration being over the infinite plane. Similarly, the S_2 stream is given momentum at the rate

$$\mu_0\iint(H_2^2 + H_1H_2\cos\theta)ds \quad . \quad . \quad . \quad (43)$$

so that the total rate of change of momentum at the end wire is

$$\mu_0\iint H^2ds = L_0I^2 \quad . \quad . \quad . \quad . \quad (44)$$

where L_0 is the inductance of the circuit per unit length of the transmission line.

(4.3) The Force on the End Wire

Suppose the end wire to move a very small distance δx, thus lengthening the circuit by the same amount. The work done by the change of momentum is then $L_0I^2\delta x$, and this is precisely the amount of additional energy which must be provided by the source if the current is to be kept constant. We then have:

Work done by change of momentum = Mechanical work done + the increment in magnetic energy,

or $$L_0I^2\delta x = F\delta x + \tfrac{1}{2}L_0I^2\delta x$$

and $$F = \tfrac{1}{2}L_0I^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (45)$$

as before.

We shall see later that the electromagnetic momentum can be identified with the momentum of the conduction electrons, so according to our hypothesis the force on the end wire in a very long rectangular circuit is entirely due to electron inertia. The force between the two side wires cannot, of course, be explained in this way, except in a short region near the corners of the circuit. The force between two long parallel current-carrying conductors may be regarded as arising from a very small unbalance of the mutual forces between the electric charges in the wires, some of which are moving and some of which are stationary.

It is evident that this hypothesis of moving energy-mass leads directly to the conclusion that Newton's third law (equality of action and reaction) does not apply to the mutual forces of two individual current elements unless the elements are parallel to each other. The hypothesis thus gives a physical justification to the accepted classical form for the mutual force of two current elements. Ampère's law of force between two current elements, upon which Weber's electromagnetic theory was based, obeys Newton's third law whatever the direction of the current elements may be. Ampère's theory thus treats the problem from the principles of statics rather than of dynamics, and for this reason theories based upon it fail to lead to electromagnetic radiation or the equivalence of mass and energy.

(4.4) The Resultant Flux of Energy

The resultant energy flux towards the end wire DA is, vectorially,

$$S = S_1 + S_2 = (E_1 + E_2) \times H \quad . \quad . \quad . \quad (46)$$

Since each side wire contains equal amounts of stationary positive charge and the wires have the same section, it follows that

$$\iint \{(E_3 + E_4) \times H\} \,.\, dS = 0 \quad . \quad . \quad . \quad (47)$$

for we may divide this integral into two equal and opposite halves by a plane half way between the side wires and perpendicular to the plane of the circuit. The resultant total rate of energy flow towards the end wire is therefore

$$P = \iint \{(E_1 + E_2 + E_3 + E_4) \times H\} \,.\, dS = \iint (E \times H) \,.\, dS \quad . \quad . \quad . \quad . \quad (48)$$

That is, it is the same as that given by the classical Poynting vector $E \times H$. This energy is converted into heat in the end wire.

Any change in the resultant electromagnetic field at a point is effected in accord with classical electromagnetic theory, with the velocity of an electromagnetic wave, which in free space is c. For example, consider Fig. 5, which shows conditions near a straight conductor which carries a changing current i. The main energy flux, $S_v = E_n \times H$, travels with velocity v parallel to the wire in the direction of electron flow. If the current increases at the rate di/dt, an induced electric field $E_i = -dA/dt$ is produced parallel to the wire in a direction opposite to that of the current. The classical Poynting vector $S_c = E_i \times H$ is then directed radially out from the wire, and represents a flow of electromagnetic energy moving outwards with velocity c. This increases the classical field energy $\frac{1}{2}(HB + ED)$, and may therefore be considered as providing the increment in the kinetic energy of the accelerating main energy flux. If the current decreases, E_i and S_c reverse, and the main energy flux decelerates.

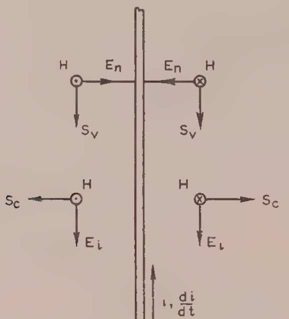


Fig. 5.—Energy fluxes when the current is changing.

(5) THE ELECTROMAGNETIC MASS AND CONDUCTION CHARGE OF A CURRENT CIRCUIT

If we attempt to apply the above detailed analysis of moving energy streams, whose energy density at a point is specified, to circuits of any shape it is clear that we shall encounter great difficulties; for each current element contributes its own component of energy flux, parallel to itself, at a given point, and each of these components must be considered to exist independently. We shall therefore proceed on the assumption that, in general, the total kinetic energy of the moving energy mass is equal to the magnetic energy of the circuit $\frac{1}{2}LI^2$.

If the mean velocity of the conduction electrons in a complete current circuit can be taken to have the same value, v, at all parts of the circuit, we can then introduce the concept of the electromagnetic mass of the circuit, which we may call M_0 when $v \ll c$. We define M_0 by the relation

$$\tfrac{1}{2}M_0v^2 = \tfrac{1}{2}LI^2, \text{ or } M_0 = L\left(\frac{I}{v}\right)^2 \quad . \quad . \quad . \quad (49)$$

L being the inductance of the complete circuit. The total electromagnetic momentum is then

$$p_{total} = M_0v = LI^2/v \quad . \quad . \quad . \quad . \quad . \quad (50)$$

If ρ is the charge of the conduction electrons comprising the current per unit length of wire, $I = \rho v$ and

$$M_0 = L\rho^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (51)$$

We now take the radical step of identifying M_0 with the total mass of the conduction electrons whose motion comprises the current. We have already noted that the identification of the magnetic energy of a free electron with its kinetic energy is generally accepted, so it is no more than a logical extension of this idea to postulate the same for the electrons in a complete current circuit. Indeed, a failure to do so seems to leave an untidy inconsistency in electromagnetism, however minute the practical consequences of this inconsistency may be. By making this identification, moreover, we can bring the known effects of electron inertia in a circuit into the general electromagnetic scheme—a unification which seems highly desirable.

Let the charge ρ consist of N conduction electrons, per unit length of wire, each of mass m_0, so that $Ne = \rho$, where $e \; (< 0)$ is the electronic charge. If l is the total length of the wire, the total momentum of the conduction electrons is Nlm_0v, and equating this to M_0v gives

$$M_0 = Nlm_0 = \frac{\rho l m_0}{e} \quad . \quad . \quad . \quad . \quad (52)$$

So from eqn. (51) we obtain

$$\rho = -\sqrt{\frac{M_0}{L}} = \left(\frac{m_0}{e}\right)\frac{l}{L} \quad . \quad . \quad . \quad . \quad (53)$$

$$M_0 = \left(\frac{m_0}{e}\right)^2\frac{l^2}{L} \quad . \quad . \quad . \quad . \quad . \quad (54)$$

and

$$\frac{M_0}{\rho} = -\sqrt{(LM_0)} = \left(\frac{m_0}{e}\right)l \quad . \quad . \quad . \quad (55)$$

We also have

$$v = \frac{I}{\rho} = -I\sqrt{\frac{L}{M_0}} = \left(\frac{e}{m_0}\right)\frac{LI}{l} \quad . \quad . \quad . \quad (56)$$

v and I being in opposite directions, since e is negative. This may be put in the form

Magnetic flux-linkages per unit length of wire, LI/l = Mean momentum of a current electron, divided by the electronic charge, m_0v/e . . . . (57)

(6) THE CONDUCTION CHARGE AND ELECTRON FLOW

From the last Section we see that, if the identity of M_0 with Nlm_0 is to be valid, the mean velocity of the conduction electrons, assumed here to be small compared with c, must be determined both by the current and the geometry of the circuit. The charge ρ is also determined, or quantized, by the geometry of the circuit. This may at first sight appear to be a difficult condition. A current in a wire is often taken to be due to the motion of all the conduction electrons in the atomic structure (of the order of 10^{23} per cubic centimetre for copper) drifting along the wire with a very low velocity which depends solely upon the resultant electric field in the conductor. This is not, however, a necessary interpretation of the theory of electronic conduction. Calculations[8] based on the Fermi–Dirac statistics, on the assumption of one conduction electron per atom, give the r.m.s. velocity of a free electron in copper as about $1{\cdot}6 \times 10^6$ m/sec. This means that individual electrons have velocities somewhere within a range which extends far below and far above this figure. If we resolve the velocity of every electron along the wire, when no current is flowing, there will be equal streams of electrons in opposite directions, with a great range of velocities in each stream. When a current flows, the above hypothesis requires that it must be provided by the requisite number of electrons whose longitudinal velocities are as close as possible to the value given by eqn. (56). This condition is satisfied if the opposing streams of electrons possessing this particular mean velocity are no longer equal, so that the difference in their charge per unit length is equal to ρ. Provided that this theory does not lead to absurd results, such as a value for ρ greater than the total charge available, there seems to be no *a priori* objection to it.

(7) TWO PRACTICAL CASES

(*a*) *A concentric cable, of inductance* 5×10^{-4} *henry/km.*

The inductance per metre length of single conductor is

$$\frac{L}{l} = 2{\cdot}5 \times 10^{-7} \text{ henry/metre;}$$

$$\frac{e}{m_0} = -\frac{1{\cdot}6 \quad \times 10^{-19}}{9{\cdot}11 \times 10^{-31}} = -1{\cdot}76 \times 10^{11}$$

so that

$$v = \left(\frac{e}{m_0}\right)\frac{L}{l}I = -4{\cdot}4 \times 10^4 I \text{ metres/sec}$$

where I is the current in amperes. The moving charge per unit length is

$$\rho = \frac{I}{v} = -2{\cdot}27 \times 10^{-5} \text{ coulomb/metre}$$

and $\quad M_0 = L\rho^2 = 2{\cdot}57 \times 10^{-13}$ kilogramme/kilometre.

(*b*) *A toroidal coil, of ring radius* $R = 10^{-1}$ *m, mean turn radius* $r = 2 \times 10^{-2}$ *m, with* $N = 1\,000$ *turns and carrying a current of* 5 *amp.*

The inductance is $L \simeq \mu_0 r^2 N^2/2R = 8\pi \times 10^{-4}$ henry, and the internal flux density is about 10^{-2} weber/m² or 100 gauss. The length of wire is $l = 2\pi rN = 40\pi$ metres, so

$$v = \left(\frac{e}{m_0}\right)\frac{L}{l}I = -(3{\cdot}52 \times 10^6)I$$
$$= 1{\cdot}76 \times 10^7 \text{ metres/second} = 0{\cdot}0585c$$

$$\rho = I/v \quad = -2{\cdot}82 \times 10^{-7} \text{ coulomb/metre}$$

$$M_0 = L\rho^2 = \; 2{\cdot}012 \times 10^{-16} \text{ kilogramme}$$

The effective number of moving conduction electrons per metre of wire is

$$N = \rho/e = 1{\cdot}76 \times 10^{12}$$

Suppose the toroid to be wound with two layers of 500 turns each, with wire of diameter 1 mm. Then the volume of the conductor per metre length is $\pi/4\,\text{cm}^3$ and the effective number of conduction electrons per cubic centimetre is $2{\cdot}24 \times 10^{12}$. This is extremely small in comparison with the total available number of conduction electrons, which is of the order of 10^{23} per cubic centimetre.

(8) SELF-INDUCED E.M.F.

It is evident that the e.m.f. induced in a circuit when the current changes, according to this hypothesis, is merely an effect of electron inertia. If the conductor has no longitudinal freedom of movement this e.m.f., in the direction of the current I, is given by

$$V_e = -L\frac{dI}{dt} = -\frac{\frac{d}{dt}(\frac{1}{2}LI^2)}{I} = -\frac{\frac{d}{dt}(\frac{1}{2}M_0v^2)}{\rho v}$$

which from eqns. (53) and (54) gives

$$V_e = \sqrt{(LM_0)}\frac{dv}{dt} = -\left(\frac{m_0}{e}\right)l\frac{dv}{dt} \quad . \quad . \quad (58)$$

or, if the velocity v is not the same at all parts of the circuit,

$$V_e = -\frac{m_0}{e}\oint\frac{dv}{dt}dl \quad . \quad . \quad . \quad . \quad (59)$$

The equivalent self-induced electric-field intensity, of classical theory, in the conductor is evidently

$$E = -\frac{m_0}{e}\frac{dv}{dt} \quad . \quad . \quad . \quad . \quad . \quad . \quad (60)$$

so that

$$eE = -m_0\frac{dv}{dt} \quad . \quad . \quad . \quad . \quad . \quad (61)$$

which may be compared with eqn. (57).

The classical law of conduction, $\boldsymbol{J} = \boldsymbol{E}/r$, where $\boldsymbol{J}$ is the current density and r the resistivity, states that the current density is, at every instant, proportional to the resultant electric-field intensity. That is, $\boldsymbol{E}$ is regarded as the cause of the current, whereas by the present hypothesis a current continues to flow in a short-circuited circuit because of the inertia of the conduction electrons. Classical theory requires this self-induced $\boldsymbol{E}$ because electron inertia is considered to be negligible or absent. The average force opposing the motion of the conduction electrons is $-\boldsymbol{J}r$ per unit charge. The electrons are decelerated by this resistance to motion, and in classical theory their deceleration induces an electric field $\boldsymbol{E}$ which cancels the decelerating force. The resultant force acting on the electron, regarded as a particle without inertia, is therefore zero. So if we use our present hypothesis in a theory of conduction, we must not include a self-induced electric-field intensity.

The linking magnetic flux, if v is single-valued, is

$$\Phi = LI = -v\sqrt{(LM_0)} = \left(\frac{m_0}{e}\right)lv \quad . \quad . \quad (62)$$

or, in general,

$$\Phi = \oint\left(\frac{m_0}{e}\right)vdl \quad . \quad . \quad . \quad . \quad . \quad (63)$$

But since

$$\Phi = \oint \boldsymbol{A} \, . \, \boldsymbol{dl}$$

where $\boldsymbol{A}$ is the magnetic vector potential at the axis of the wire, we can put

$$\boldsymbol{A} = \left(\frac{m_0}{e}\right)\boldsymbol{v} + \text{grad}\,\psi \quad . \quad . \quad . \quad . \quad (64)$$

when $\boldsymbol{A}$ is due to the current in the conductor alone, ψ being an arbitrary scalar potential function.

(9) VARIATION OF v IN A CONDUCTOR OF FINITE CROSS-SECTION*

We have tacitly assumed that the conductor has a negligibly small cross-section, so that the electron velocity given by eqn. (56)

* The case of a finite conductor is treated more rigorously in the Appendix.

applies to all parts of the section. Actually v must vary across a conductor of finite section, for from eqn. (64) we have

$$\operatorname{curl} \boldsymbol{v} = \frac{e}{m_0} \operatorname{curl} \boldsymbol{A} = \frac{e}{m_0} \boldsymbol{B} \quad . \quad . \quad . \quad . \quad (65)$$

where $\boldsymbol{B}$ is the flux density inside the conductor. From eqn. (60) this is seen to be consistent with the classical equation curl $\boldsymbol{E} = -\dot{\boldsymbol{B}}$.

(10) RELATIVISTIC EXPRESSIONS FOR THE MAGNETIC ENERGY AND THE SELF-INDUCED E.M.F.

So far we have assumed that the velocity v of the effective conduction electrons is sufficiently small in comparison with c to justify the neglect of any variation, with velocity, of the electromagnetic mass. In the example of the toroidal coil, however, we obtained an electron velocity of about $c/17$. It is thus evident that, in circuits of high L/l with high currents, the variation of mass may be significant.

If v is not negligible in comparison with c the kinetic or magnetic energy is not $\frac{1}{2}M_0v^2$ but

$$U_m = (M - M/\beta)c^2 \quad . \quad . \quad . \quad . \quad . \quad (66)$$

where M is the electromagnetic mass of the moving energy and M/β is its rest mass, not necessarily equal to M_0. For simplicity we assume that the circuit is such that v is the same over its entire length, for otherwise we shall encounter great complexity. We require U_m as a function of the current, and proceed on the assumption that the relation between electron momentum and flux linkage, eqn. (57), remains unchanged. Since the electronic mass is now $m = \beta m_0$, this requires

$$\rho = \frac{\beta m_0}{e} \frac{l}{L} = \beta \rho_0 \quad . \quad . \quad . \quad . \quad . \quad (67)$$

so that the number of conduction electrons per unit length is $N = \beta N_0$, where ρ_0 and N_0 are the low-velocity values.

The electromagnetic mass is then

$$M = Nlm = \beta^2 M_0 \quad . \quad . \quad . \quad . \quad . \quad (68)$$

We also have $$v = \frac{I}{\rho} = -\frac{I}{\beta}\sqrt{\frac{L}{M_0}} = -I\sqrt{\frac{L}{M}} \quad . \quad . \quad . \quad (69)$$

whence $$v^2 = \frac{LI^2}{M_0}\left(1 + \frac{LI^2}{M_0c^2}\right)^{-1} \quad . \quad . \quad . \quad . \quad (70)$$

If $LI^2 = M_0c^2$, then $I = m_0lc/eL \geqslant 1{\cdot}7 \times 10^{-3}l/L$. Denote the latter quantity by a. Then

$$\left.\begin{array}{ll} (a) \text{ if } I \ll a, & v = -I\sqrt{\dfrac{L}{M_0}} \\ (b) \text{ if } I \gg a, & v \to c \end{array}\right\} \quad . \quad . \quad . \quad (71)$$

From eqn. (70) we obtain

$$\beta = \left(1 + \frac{LI^2}{M_0c^2}\right)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad (72)$$

and from eqns. (66) and (68)

$$U_m = \beta M_0 c^2(\beta - 1) \quad . \quad . \quad . \quad . \quad . \quad (73)$$

This reduces to $\frac{1}{2}LI^2$ if $I \ll a$, but approaches LI^2 if $I \gg a$.

The self-induced e.m.f., V_e, is given by $V_eI = -dU_m/dt$, whence

$$V_e = -L\frac{dI}{dt}\left\{2 - \left(1 + \frac{LI^2}{M_0c^2}\right)^{-1/2}\right\} \quad . \quad . \quad (74)$$

which reduces to $-L\dfrac{dI}{dt}$ if $I \ll a$, and approaches $-2L\dfrac{dI}{dt}$ if $I \gg a$.

For the toroidal coil (Section 7), $l/L = 5 \times 10^4$ and $a \geqslant 85$ amp. This is greatly in excess of the current-carrying capacity of the wire. Our theory thus becomes non-linear under conditions which are outside the range of normal practice, and this departure from classical electromagnetism must also, of course, apply to the forces experienced by current-carrying conductors.

As will be noted in the next Section, Maxwell developed his general electromagnetic theory on the assumption that the carriers of current in a conductor have no inertia. A study of Chapters 5–9 of the fourth part of his *Treatise* will show how classical theory is deeply rooted in the general principles of dynamics, and Maxwell certainly regarded magnetic energy as *kinetic*. He called the vector potential $\boldsymbol{A}$ the *electrokinetic momentum*, but without attempting to associate with it any particular velocity. Now it is impossible to conceive the ideas of kinetic energy and momentum without accepting the concept of moving mass, which in Maxwell's day was uncomplicated by the equivalence of mass and energy and the variation of mass with velocity. It should not therefore surprise us if our present hypothesis throws doubt on the complete validity of Maxwell's equations.

(11) ELECTRON-INERTIA EFFECTS

(11.1) History and Present Theoretical Position

In his "Treatise on Electricity and Magnetism" (Part IV, Chapter VI), Clerk Maxwell discussed three types of experimental effect which should exist if an electric current in a conductor is a true motion of some substance having inertia.

(*a*) If a circular coil is freely suspended by an axial thread with its axis vertical, any change in the current flowing in it should be accompanied by a rotation of the coil.

(*b*) A coil carrying current should exhibit gyroscopic effects and, if Ampère's hypothesis that ferromagnetism is due to atomic currents is correct, the same should apply to a magnet.

(*c*) When a rapidly rotating coil, which is part of an unenergized closed conducting circuit, is suddenly stopped, the inertia of the current carriers should cause a momentary displacement of electricity (i.e. a current) through the circuit.

Maxwell stated that no such phenomena had ever been observed, and apparently performed an experiment to test (*b*), but without a positive result. He showed that all three effects depend on the possible existence in the expression for the kinetic energy of a moving current-carrying conductor of a term involving the product of the velocity of the conductor and the velocity of the electricity relative to it. He called this term T_{me}, and concluded:

> We have thus three methods of detecting the existence of the terms of the form T_{me}, none of which have hitherto led to any positive result. I have pointed them out with the greater care because it appears to me important that we should attain the greatest amount of certitude within our reach on a point bearing so strongly on the true nature of electricity.

He therefore developed his electromagnetic theory on the assumption that such effects do not exist, or at least that they produce no sensible effect, and they cannot be deduced from his fundamental equations of the electromagnetic field.

Nevertheless, all three types of effect have now been experimentally observed.[9] The first successful experiments on the gyromagnetic effect were those of Barnett[10] in 1915, who succeeded in magnetizing an iron rod by rotating it. The converse effect, the production of rotation by magnetization, was observed by Einstein and de Haas[11–13] (1915 and 1916). A conclusion from these and later experiments is that the magnetic moment of a ferromagnetic atom must be due to spinning electrons rather than to orbital electrons. In 1916 Tolman and Stewart[14] succeeded in detecting effect (*c*). Furthermore, their experiments proved that the carriers of electricity in a conduction current have a negative charge, and the ratio e/m for a conduction electron was found to be approximately the same as that for a free electron. Effect (*a*) was sought by Sir Oliver Lodge,[16] but was not detected until 1930, when experiments by Barnett[17] gave positive and theoretically consistent results. Barnett used alternating current, tuning the frequency so that mechanical resonance occurred, and measured the torque by removing the oscillations by means of an equal and opposite torque which could be calculated.

In the past treatment of electron-inertia effects it has been usual to take the momentum and kinetic energy of the conduction electrons as extra-electromagnetic phenomena, ρ being taken as equal to the charge of the total number of conduction electrons available, so that the mean velocity v becomes very small. For a given current, so that ρv is constant, the kinetic energy of the electrons is proportional to v, and this older viewpoint gives a kinetic energy which is extremely small in comparison with the magnetic energy of the circuit. It has therefore previously been held that it is not possible to identify the two energies, and the total energy in a stationary circuit due to the current has been taken as being equal to the magnetic energy plus a very small correction for the kinetic energy of the conduction electrons.[18] According to our hypothesis, however, the magnetic energy of the circuit and the kinetic energy of the conduction electrons are the same thing, and in a stationary current circuit the energy due to the current flow is exactly equal to either. It also follows from this theory that these very small manifestations of electron inertia in a current circuit are a necessary consequence of the momenta of the energy fluxes. We shall apply our hypothesis to the experiment of Tolman and Stewart [effect (*c*)] and to the Barnett effect (*a*). Experiments on the gyromagnetic effect have been carried out on iron rods, which are outside the scope of the paper.

(11.2) The Experiment of Tolman and Stewart: Production of an Inertial Current

A rapidly rotating coil is connected, by sliding contacts, in a stationary circuit which includes a ballistic galvanometer. The circuit contains no source of e.m.f., so that when the coil is at rest or rotating uniformly there is no current. When it is suddenly stopped the momentum of the conduction electrons carries them on and the galvanometer registers the charge displaced around the circuit.

Consider the coil to be rotating uniformly, with no current, with peripheral velocity w. As before, let there be N conduction electrons per unit length of wire, so that $\rho = Ne$. The momentum of these electrons, due to the rotation, is $Nmwl$ where m is the mass of an electron. If I is the current which would have the same energy-stream momentum, we have

$$LI^2/v = L\rho^2 v = Nmwl$$

and

$$I = \rho v = \frac{Nmwl}{L\rho} = \frac{mwl}{Le} \quad . \; . \; . \; . \; . \quad (75)$$

The quantity of electricity displaced when the rotation is stopped will be the same as that displaced when a current of this value is dissipated in a circuit of inductance L and resistance R. Since in such a case $L\frac{di}{dt} + Ri = 0$,

$$L\int_I^0 di + R\int_I^0 idt = 0$$

so that

$$LI - RQ \text{ and } Q = \frac{LI}{R} = \frac{mwl}{Re}$$

and

$$\frac{e}{m} = \frac{wl}{RQ} \quad . \; . \; . \; . \; . \; . \; . \quad (76)$$

This is the same relation as used by Tolman and Stewart, and their experiment gave results consistent with a value of e/m approximately equal to that for a slowly moving free electron. More recent experiments by Kettering and Scott[19] have confirmed this identity to a considerable degree of accuracy.

(11.3) The Barnett Effect

The experiment of Barnett,[17] in 1930, confirmed that when the current in a circuit changes, the conductor experiences a very small longitudinal force. Suppose the circuit includes a helical coil free to rotate about its axis, which is vertical. If the coil is stationary and carries a current I, the angular momentum of the energy stream is M_0vr, where r is the mean radius of the coil and M_0 is the electromagnetic mass of the coil. Let the mass of the wire be M_c; then when the current is stopped the electromagnetic angular momentum M_0vr is converted into mechanical angular momentum M_cwr, where w is the final peripheral velocity of the wire. Thus

$$w = \frac{M_0}{M_c}v \quad . \; . \; . \; . \; . \; . \; . \quad (77)$$

Let us apply this to the case of a solenoid of the same axial length, turn area and number of turns as the toroid previously considered. For a rough calculation of the order of magnitude of the effect we may take the inductance, and hence M_0, to be the same as that of the toroid. The mass of the wire will be about 0·9 kg, so that, if a current of 5 amp is stopped, the final peripheral velocity of the wire will be

$$w \simeq \frac{2 \times 10^{-16} \times 1{\cdot}76 \times 10^7}{0{\cdot}9} \simeq 4 \times 10^{-9} \text{ metre/second}$$

in the same direction as the original electron flow, and opposed to the direction of the original current.

(11.4) The Self-Induced E.M.F.

Since classical electromagnetic theory is based on Maxwell's assumption that the kinetic energy of a moving current circuit is independent of the velocity product vw, and since the existence of these electron-inertia effects shows that this is not rigorously true, it follows that the fundamental law of self-induction, when a circuit is not constrained by external mechanical forces, is no more than a very close approximation.

Consider the freely suspended solenoid, initially stationary and carrying a steady current. If the circuit is short-circuited the energy momentum will be transferred to the coil as a whole. If the total mass of the wire, M_c, is taken as including the electromagnetic mass M_0, and M_0 is taken as constant, the total kinetic energy during the acceleration of the coil is given by

$$\begin{aligned}\text{Kinetic energy} &= \tfrac{1}{2}(M_c - M_0)w^2 + \tfrac{1}{2}M_0(w+v)^2 \\ &= \tfrac{1}{2}M_cw^2 + M_0wv + \tfrac{1}{2}LI^2 \quad . \; . \; . \quad (78)\end{aligned}$$

the middle term being Maxwell's T_{me}. The total momentum is constant, so that

$$(M_c - M_0)\frac{dw}{dt} + M_0\left(\frac{dw}{dt} + \frac{dv}{dt}\right) = 0$$

and

$$M_c\frac{dw}{dt} = -M_0\frac{dv}{dt} \quad . \; . \; . \; . \; . \; . \quad (79)$$

Then since there are no external forces, we have

Rate of increase of kinetic energy + Rate of energy conversion into heat = 0, and if V_e is the self-induced e.m.f., it follows that

$$M_c w\frac{dw}{dt} + M_0\left(w\frac{dv}{dt} + v\frac{dw}{dt}\right) + LI\frac{dI}{dt} + V_e I = 0 \ . \quad (80)$$

whence from eqn. (79),

$$V_e = -L\frac{dI}{dt} - \frac{M_0 v}{I}\frac{dw}{dt}$$

$$= -\frac{dI}{dt}\left[L - \frac{1}{M_c}\left(\frac{M_0}{\rho}\right)^2\right]$$

$$= -L\left(1 - \frac{M_0}{M_c}\right)\frac{dI}{dt} \quad . \quad . \quad . \quad . \quad . \quad (81)$$

from eqn. (51). If, however, the circuit is constrained by external forces in such a way that the conductors have no longitudinal freedom of movement, the induced e.m.f. will suffer no modification.

For the given solenoid, $M_0/M_c \simeq 2{\cdot}3 \times 10^{-16}$. The original magnetic energy of the stationary solenoid is $\frac{1}{2}LI^2 = 10^{-2}\pi$ joule, and the final energy of the rotating coil, assuming a perfectly free suspension and no frictional loss, is $\frac{1}{2}M_c w^2 = 7{\cdot}2 \times 10^{-18}$ joule. Except for this extremely small amount of energy, all the original magnetic energy is converted into heat in the conductor.

(11.5) The Force on the Conductor

The longitudinal force on the conductor, when the current changes, is $-M_0 d(v + w)/dt$, but since $w \ll v$

$$F \simeq -M_0\frac{dv}{dt} = -\frac{M_0}{\rho}\frac{dI}{dt} = -\left(\frac{m_0}{e}l\right)\frac{du}{dt}$$

$$= -\frac{m_0}{e}\frac{dI}{dt} \text{ per unit length of wire} \quad . \quad . \quad (82)$$

If the resistance of the solenoid is 2·72 ohms, the maximum value of $-dI/dt$ when a current of 5 amp is short-circuited is $IR/L = 5{\cdot}4 \times 10^3$ amp/sec. The maximum value of the longitudinal force is therefore about 3×10^{-8} newton/m, or 3×10^{-5} dyne per centimetre of wire.

(12) THE E.M.F. INDUCED BY THE LONGITUDINAL ACCELERATION OF A CONDUCTOR

Whenever the longitudinal velocity of a conductor changes, there is a very small induced e.m.f., as shown in Section 11.4. If the longitudinal acceleration is du/dt, in order to prevent an induced current the electromagnetic mass M_0 must be given the same acceleration, and this requires an impressed force, per unit charge, equivalent to an electric field of intensity $(M_0/\rho l)du/dt$. The impressed e.m.f. required to prevent the induced current is therefore $V_e' = (M_0/\rho)du/dt$, and the e.m.f. induced in the wire by the acceleration is equal and opposite to this, i.e.

$$V_e = -\frac{M_0}{\rho}\frac{du}{dt} = \sqrt{(LM_0)}\frac{du}{dt} = -\left(\frac{m_0}{e}l\right)\frac{du}{dt} \quad . \quad (83)$$

In M.K.S. units

$$V_e = 5{\cdot}7 \times 10^{-12} l\frac{du}{dt} \text{ volts} \quad . \quad . \quad . \quad . \quad . \quad (84)$$

in the direction of the acceleration. This e.m.f. is additional to that given by the classical laws of electromagnetic induction.

(13) INERTIAL CURRENT IN A PERFECT CONDUCTOR

If a perfectly conducting ring, without current, is rotating with a peripheral velocity u and is then stopped, a steady "super-current" should be produced, since the conduction charge ρ should continue to move with velocity u. This current should therefore be

$$I = \rho u = \left(\frac{m_0}{e}\right)\frac{l}{L}u \quad . \quad . \quad . \quad . \quad . \quad (85)$$

producing a linking flux

$$\Phi = IL = \left(\frac{m_0}{e}\right)lu \quad . \quad . \quad . \quad . \quad (86)$$

The mean flux density through the ring, if R is the mean radius and ω the angular velocity, is

$$B = \frac{\Phi}{\pi R^2} = 2\left(\frac{m_0}{e}\right)\omega \quad . \quad . \quad . \quad . \quad (87)$$

It is of interest to note that this is the same as the flux density inside a rotating superconducting sphere as deduced from the London electrodynamic theory of superconductivity.[20] By the London theory, however, the current and field should be independent of the previous deceleration, i.e. if a ring with no current is rotating with constant speed at a normal temperature and is then supercooled, the supercooling alone should generate the supercurrent. This curious result does not follow from our theory. Although the expected effect is very small, it should be possible to determine the truth by experiment.

(14) THE ENERGY REQUIRED TO ESTABLISH A CURRENT IN A COIL ROTATING AT CONSTANT SPEED

Consider a circular coil rotating about its axis with constant angular velocity, so that the linear velocity of the conductor is u, taken positive in the direction of the current (Fig. 6). Then the total kinetic energy due to the current flow is

$$\tfrac{1}{2}M_0[(v - u)^2 - u^2] = \tfrac{1}{2}LI^2 - M_0 uv \quad . \quad . \quad (88)$$

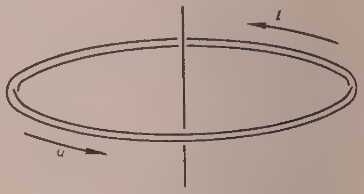


Fig. 6.—Rotating current-loop.

In order to keep the velocity of the wire constant during the growth of the current, a mechanical force must be exerted on the wire of value $M_0 dv/dt$ in the direction of v, or $-M_0 dv/dt$ in the direction of I. Thus, in raising the velocity of the electromagnetic mass M_0 relative to the wire from zero to v, mechanical work must be done by the driving motor at the rate $-M_0 u dv/dt$, and the total mechanical work required is $-M_0 uv$. The remainder of the required energy given by eqn. (88) must be provided by the source of e.m.f. causing the current. Thus the energy required to establish the current, apart from ohmic loss, is $\frac{1}{2}LI^2$ provided electrically, and $-M_0 uv = -uI\sqrt{(LM_0)}$, provided mechanically (both v and I are here taken as positive quantities). The latter term becomes positive if u is in the opposite direction to the current. The e.m.f. of self-induction is $-LdI/dt$ and is independent of u.

(15) THE LONDON THEORY OF SUPERCONDUCTIVITY

At present the macroscopic theory of superconductivity, originally due to H. and F. London,[21] is an *ad hoc* modification of classical theory, an essential difference being that account is taken of the inertia of the "superconducting" electrons. Classical theory, as we have noted, neglects electron inertia. The London theory has some points of similarity with the hypothesis presented in this paper. For example, it leads to the conclusion that "supercurrents" are quantized in a macroscopic way, for a conductor as a whole,[22] and that for a large superconductor the kinetic-energy density of the superconducting electrons is equal to their magnetic-energy density.[23] However, these two energies are regarded as being distinct and additive, and not, as we have postulated, different aspects of the same thing.

It is difficult to reconcile the London equations with some of the basic relations of electromagnetic theory. For example, a fundamental equation of the theory is our eqn. (64), but with reversed sign.[21] Now the vector potential A, where curl $\mathbf{A} = \mathbf{B}$, is parallel, not anti-parallel, to a normal current in a straight conductor, so that by the London theory the induced electric field when a supercurrent changes is quite different from that induced by the same rate of change of a similar normal current. From the form of the London equations it would appear that such peculiar and rather incredible features of the theory may be related to the inclusion of both electron inertia and the equivalent self-induced electric field of classical non-inertial theory, contrary to our conclusion in Section (8). If our present hypothesis should lead to an alternative theory of superconductivity, one would expect it to cover both superconduction and normal conduction, with superconductors appearing merely as perfect conductors ($r = 0$), but further discussion of this interesting problem is clearly beyond the scope of the paper.

(16) REFERENCES

(1) MATHUR, S. B. L.: "Biot–Savart Law and Newton's Third Law of Motion," *Philosophical Magazine* (7th Series), 1941, **32**, p. 171.

(2) DUNTON, W. F.: "Validity of Laws of Electrodynamics," *Nature*, 1937, **140**, p. 245.

(3) CLAYTON, A. E.: Discussion on the above, *ibid.*, p. 246.

(4) DUNTON, W. F., and DRYSDALE, C. V.: "A Comprehensive Fundamental Electrical Formula," *ibid.*, 1939, **143**, p. 601.

(5) HOWE, G. W. O.: "The Application of Newton's Third Law to an Electric Circuit," *Wireless Engineer*, 1945, **22**, p. 521.

(6) HOWE, G. W. O.: "Mechanical Force on the Short Side of a Long Rectangular Circuit," *ibid.*, 1952, **29**, p. 83.

(7) MOULLIN, E. B.: Discussion on the above, *ibid.*, p. 193.

(8) MOTT, N. F., and JONES, H.: "The Theory of the Properties of Metals and Alloys" (University Press, Oxford, 1936), p. 268 and Chapter 5.

(9) BARNETT, S. J.: "Gyromagnetic and Electron-Inertia Effects," *Reviews of Modern Physics*, 1935, **7**, p. 129. (For a comprehensive review.)

(10) BARNETT, S. J.: "Magnetization by Rotation," *Physical Review*, 1915, **6**, p. 239.

(11) EINSTEIN, A., and DE HAAS, W. J.: "Experimenteller Nachweis der Ampèreschen Molekularströme," *Verhandlungen der Deutschen Physikalischen Gesellschaft*, 1915, **17**, p. 152.

(12) EINSTEIN, A.: "Ein einfaches Experiment zum Nachweis der Ampèreschen Molekularströme," *ibid.*, 1916, **18**, p. 173.

(13) DE HAAS, W. J.: "Weitere Versuche über die Realität der Ampèreschen Molekularströme," *ibid.*, p. 423.

(14) TOLMAN, R. C., and STEWART, T. D.: "The Electromotive Force produced by the Acceleration of Metals," *Physical Review*, 1916, **8**, p. 97.

(15) TOLMAN, R. C., and STEWART, T. D.: "The Mass of the Electric Carrier in Copper, Silver and Aluminium," *ibid.*, 1917, **9**, p. 164.

(16) LODGE, O. J.: "Modern Views of Electricity" (Macmillan, London, 1892), Second edition, p. 97.

(17) BARNETT, S. J.: "A New Electron-Inertia Effect and the Determination of m/e for the Free Electrons in Copper," *Philosophical Magazine* (7th Series), 1931, **12**, p. 349.

(18) LORENTZ, H. A.: "The Motion of Electricity in Metals," *Journal of the Institute of Metals*, 1925, **33**, p. 265. A summary appeared in *Engineering*, 1925, **119**, p. 625.

(19) KETTERING, C. F., and SCOTT, G. G.: "Inertia of the Carrier of Electricity in Copper and Aluminium," *Physical Review*, 1944, **66**, p. 257.

(20) LONDON, F.: "Superfluids. Volume 1: Macroscopic Theory of Superconductivity," (Wiley, New York, 1950), p. 82.

(21) LONDON, F., and LONDON, H.: "Electromagnetic Equations of the Supraconductor," *Proceedings of the Royal Society*, A, 1935, **149**, p. 71.

(22) LONDON, F.: "Superfluids, Volume 1," pp. 2 and 3.

(23) LONDON, F.: *ibid.*, p. 66.

(24) LIVENS, G. H.: "The Theory of Electricity" (University Press, Cambridge, 1918), p. 553.

(25) LIVENS, G. H.: *ibid.*, p. 555.

(26) SLEPIAN, J.: "Energy Flow in Electric Systems—the V_i Energy-Flow Postulate," *Transactions of the American I.E.E.*, 1942, **61**, p. 835.

(27) CARTER, G. W.: "The Electromagnetic Field in its Engineering Uses" (Longmans, London, 1954).

(28) MAXWELL, J. C.: "A Treatise on Electricity and Magnetism" (University Press, Oxford; Third Edition, 1892). Vol. II, Part IV, p. 634, eqn. (16), p. 322.

(17) APPENDIX*

In this Monograph the conventional idea of field energy has been used, and the concept of a component energy flux of the Poynting form was introduced. This led to eqn. (64) for the relation between A and v for a filamentary circuit, and in obtaining eqn. (65) it was tacitly assumed that eqn. (64) is applicable to a conductor of finite section. To justify this it is necessary to express the total magnetic energy of the current in a way which attributes it to the interior of the conductor and not throughout the whole of the magnetic field.

If T is the magnetic energy and W the total electromagnetic energy within any given volume which encloses the circuit,[24]

$$\frac{dT}{dt} = \frac{dW}{dt} - \iiint (\boldsymbol{E} \,.\, \boldsymbol{J}) d\tau \quad . \quad . \quad . \quad (89)$$

where $\boldsymbol{J}$ is the total current density, including the displacement current. It is assumed that the current is quasi-steady, so that radiation of energy can be neglected. Following Livens[25] we put $\boldsymbol{E} = -\frac{\partial A}{\partial t} - \text{grad}\,\phi$, where ϕ is the electric scalar potential, so that

$$\begin{aligned}\iiint (\boldsymbol{E} \,.\, \boldsymbol{J}) d\tau &= -\iiint \left(\frac{\partial A}{\partial t} \cdot \boldsymbol{J}\right) d\tau - \iiint (\boldsymbol{J} \,.\, \text{grad}\,\phi) d\tau \\ &= -\iiint \left(\frac{\partial A}{\partial t} \cdot \boldsymbol{J}\right) d\tau + \iiint (\phi \,\text{div}\, \boldsymbol{J}) d\tau - \iint \phi J_n dS \\ &\quad . \quad . \quad . \quad . \quad (90)\end{aligned}$$

where J_n is the component of $\boldsymbol{J}$ normal to the boundary surface of the volume. Since div $\boldsymbol{J} = 0$ we then obtain

$$\frac{dT}{dt} = \frac{dW}{dt} + \iint \phi J_n dS + \iiint \left(\frac{\partial A}{\partial t} \cdot \boldsymbol{J}\right) d\tau \ . \quad (91)$$

* The Appendix was received 18th November, 1955.

The term ϕJ_n represents an energy flux through the surface.* If it be denoted by S_n it follows that

$$-\frac{dW}{dt} = \iint S_n dS = \iint \phi J_n dS \quad . \quad . \quad . \quad (92)$$

and thus

$$T = \iiint d\tau \int_0^A \boldsymbol{J} . d\boldsymbol{A} \quad . \quad . \quad . \quad . \quad (93)\dagger$$

The displacement current is negligible inside the conductor, so if the integration is confined to the volume of the conductor we may take $\boldsymbol{J}$ in eqn. (93) as consisting entirely of conduction current. Furthermore, since there is no radiation the magnetic field of the external displacement current can also be neglected. The normal component of conduction current J_n can exist only when the surface charges, and therefore the external electric field, are changing and the outward normal component of the displacement current from the surface is equal to J_n. Thus S_n represents the outward flow of energy through the surface necessary to provide the energy increase in the external electric field. Since $\boldsymbol{E}$ inside the conductor is negligible compared with the external field, we have a consistent scheme in which, to a very close approximation, all the magnetic or kinetic energy is confined within the conductor and all the electric or potential energy is outside it.

Since $\boldsymbol{J}$ is entirely conduction current within the conductor we have

$$\boldsymbol{J} = en_e\boldsymbol{v} \quad . \quad . \quad . \quad . \quad . \quad . \quad (94)$$

where n_e is the number of effective conduction electrons per unit volume. The kinetic energy density is $\frac{1}{2}mn_e v^2 = \frac{1}{2}J(m/e)v$, m being the effective electromagnetic mass per electron. So if the magnetic energy of the current is to be equal to the kinetic energy of the effective conduction electrons we must have

$$\int_0^A \boldsymbol{J} . d\boldsymbol{A} = \frac{1}{2}J\left(\frac{m}{e}\right)v$$

or, from eqn. (94),

$$\int_0^A \boldsymbol{v} . d\boldsymbol{A} = \frac{1}{2}\left(\frac{m}{e}\right)v^2 \quad . \quad . \quad . \quad . \quad (95)$$

the solution of which is clearly

$$\boldsymbol{A} = \left(\frac{m}{e}\right)\boldsymbol{v} \quad . \quad . \quad . \quad . \quad . \quad . \quad (96)$$

Eqn. (64), with the arbitrary scalar potential function ψ taken as zero, is therefore valid for conductors of finite section, and eqn. (65) is also valid. Furthermore, these relations are not restricted to a circuit with a single-valued current and apply, for instance, to the current in a long transmission line.

* Compare the energy flux discussed by Slepian[26] and Carter.[27]

† Since $\boldsymbol{A}$ is proportional to $\boldsymbol{J}$, this result is the same as that given by Maxwell[28] as an alternative to the usual expression which assigns the electrokinetic energy to the magnetic field. He adopted the latter as being more in accord with his fundamental hypothesis of a medium or aether.

621.317.365: 621.385.1.029.6

Monograph No. 151 R
Oct. 1955

INTERPRETATION OF WAVELENGTH MEASUREMENTS ON TAPE HELICES

By C. P. ALLEN, B.Sc.(Eng.), Associate Member, and G. M. CLARKE, M.A., Ph.D., Graduate.

(*The paper was first received* 23*rd April, and in revised form* 21*st July*, 1955. *It was published as an* Institution Monograph *in October*, 1955.)

SUMMARY

During the course of an investigation of helix propagation by the measurement of the standing-wave pattern in a shielded length of helical line open-circuited at one end, it was found that the minima of the pattern were not equally spaced when the phase-change per turn approached π radians, and furthermore, this phase-change remained nearly stationary at π radians over a finite frequency band. This would indicate a highly dispersive region which is not predicted by theoretical analyses.

However, when the results are interpreted, taking into account the presence of space harmonics, an explanation of the phenomenon is obtained which indicates that there will be a region of apparent dispersion for a finite length of helix, although the basic propagation on the infinite helix is dispersionless. Such an effect may be of importance in the operation of travelling-wave devices designed to operate through this band.

(1) INTRODUCTION

The work of Sensiper[1] on unshielded tape helices has shown the presence of "forbidden" regions for propagation (shown shaded in Fig. 1) on helices which have a finite number of starts, as distinct from the sheath helix which has an infinite number of starts. This is due to the presence of space-harmonic components, all of which must decay radially away from the helix for guided propagation to take place. The interpretation suggested by Stark[2] based on reinforcement of radiation from the current elements at each turn is instructive and can be demonstrated to agree with Sensiper's forbidden regions in a simple manner.

The boundaries of the forbidden regions can be labelled on this basis, as in Fig. 1. The radiation angle is given by

$$\cos\theta = \frac{\beta_n p + 2(m-n)\pi}{\beta_0 p} \quad . \quad . \quad . \quad . \quad (1)$$

and
$$\beta_n p = \beta p + 2n\pi$$

where p is the pitch of the helix, $\beta_0 = \omega/c$, β_p is the phase-change per turn ($2\pi p/\lambda_g$), m is an integer specifying the forbidden region, and n is an integer giving the order of the radiating space harmonic. For a helix with a shielding outer conductor there is no reason why these regions should be forbidden, since the radiation is reflected back to the helix, and as suggested by Stark[2] the term "exceptional" rather than "forbidden" is more appropriate. A simple analysis similar to that given by Lines *et al.*[3] leads to the conclusion that components in the exceptional regions will increase radially towards the outer conductor, and the group velocity of the field will to the first approximation still be given by $c \sin\psi$, where ψ is the helix angle.

In the course of experimental work on the design of a tape-helix deflection system for a broad-band oscillograph, measurements were made up to a sufficiently high frequency to enter the $m = -1$ exceptional region. However, certain phenomena were first observed in the neighbourhood of the point C (see Fig. 1), and the interpretation of these will form the major subject-matter of the paper. The phenomena persisted in spite of precautions taken to reduce to negligible proportions the effect of the measuring probe on the field, and the use of a specially constructed outer conductor which enabled the slot to close on either side of the travelling probe[4] (see Figs. 2A and 2B). The same phenomena have been observed in the case of an unshielded or open helix, and instances have also been reported by Epzstein and Mourier.[5]

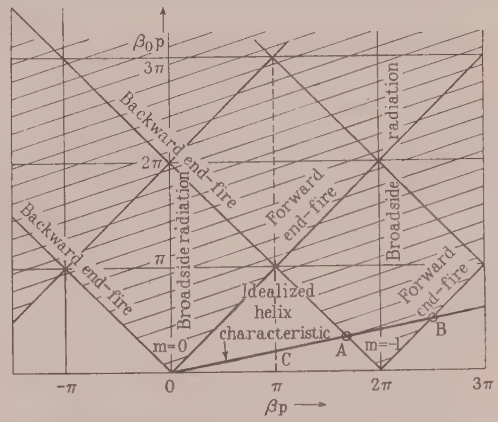


Fig. 1.—Sensiper's forbidden regions for an unshielded tape helix with fundamental of idealized helix characteristic.
Shading represents forbidden regions.

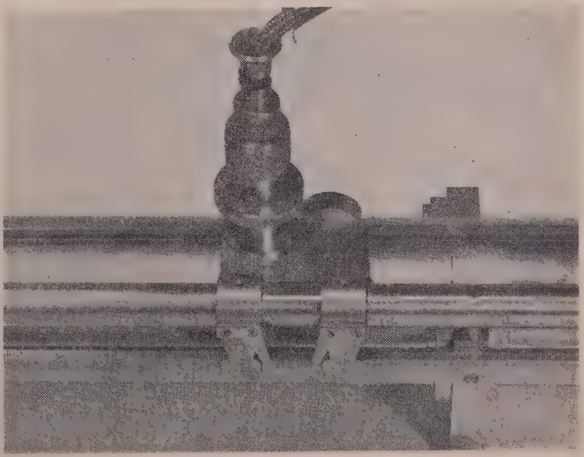


Fig. 2A.—Measuring probe and carriage, showing slot-closing device.

Correspondence on Monographs is invited for consideration with a view to publication.
The paper is a communication from the Staff of the Research Laboratories of The General Electric Company, Limited Wembley, England.

The measurements were performed with the helix open-circuited at the end remote from the generator, and the standing-wave pattern of the radial field was measured. At low frequencies the normal type of pattern was found, the maxima and minima being spaced equally along the line, and the amplitude of the maxima was constant. On approaching the point C (Fig. 1), a periodic variation in the amplitude of the maxima was observed in step with increases and decreases of the distance between adjacent minima (see Fig. 3). This effect can be accounted for by the interference between the fundamental component and the series of space harmonics travelling in the two directions. Owing to the reflective termination space-harmonic components are present corresponding to propagation of energy in the reverse direction of the line, and the phase characteristics of the fundamental and the first reverse component for both directions of energy flow for an idealized helix characteristic are shown in Fig. 4A.

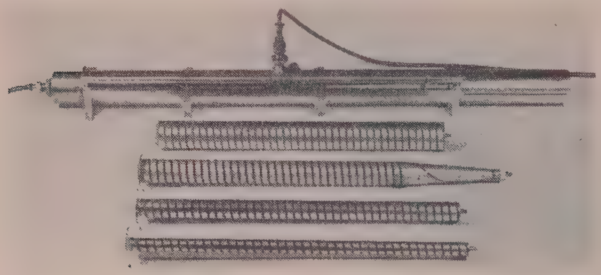


Fig. 2B.—Measuring line and some tape helices upon which experiments were performed.
Scale: one-quarter full size.

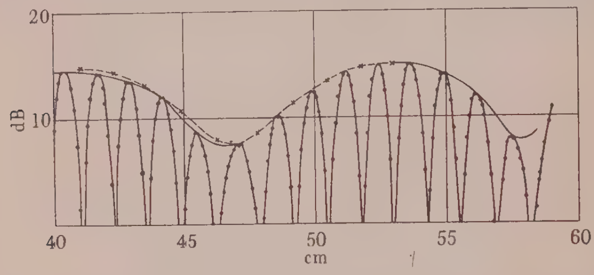


Fig. 3.—Typical pattern measured near π frequency.
– – – Theoretical curve based on a ratio of 7·5 dB maximum to minimum (fundamental and first reverse space harmonic).

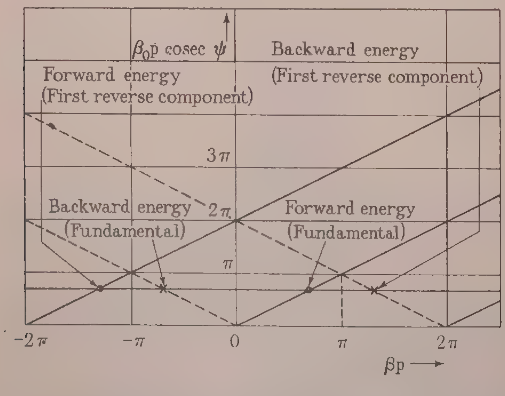


Fig. 4A.—Phase diagram showing fundamental and first reverse space harmonics in two directions.
——— Forward energy propagation.
– – – – Reflected energy propagation.

Each component for energy propagation in the forward direction can be paired with the same component in the backward direction to give a sequence of standing waves, and the interference between the members of this sequence will be discussed in the paper. It is clear that the effect will first become noticeable near the frequency giving $\beta p = \pi$, for here the amplitude and wavelength of the first reverse components approach those of the fundamentals. In fact, it has been found that only these four components are necessary to account with considerable accuracy for the experimental results in this region.

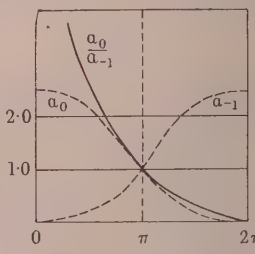


Fig. 4B.—Amplitude of fundamental and first reverse components.

It is clear that there will be a sequence of higher frequencies near which these interference phenomena will occur, the important components being the *n*th and $(n - 1)$th near to $\beta_0 p \operatorname{cosec} \psi = (n + 1)\pi$. The following discussion will be concerned only with the $n = 0$ region.

(2) DESCRIPTION OF THE EXPERIMENTAL RESULTS AND COMPARISON WITH A SIMPLE THEORY

The pattern shown in Fig. 3 is typical, and a close examination shows that the helix wavelength cannot immediately be deduced from the distance between adjacent minima, since these vary, and the envelope of the maxima is a periodic function of distance along the line.

Let the amplitude of the fundamental and first reverse components be given by a_0 and a_{-1}, respectively. Addition of the forward- and backward-travelling fundamentals gives a standing-wave pattern described by

$$a = a_0 \sin (m\theta + \alpha_0) \quad . \quad . \quad . \quad . \quad (2)$$

where θ is defined with respect to the unknown periodic envelope such that

$$\theta = \pi x/l \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where l is the distance along the helix between maximum and minimum of the envelope, and x is the axial distance. This is equivalent to stating that there are m fundamental wavelengths in a length $2l$ of the line [m is an integer unconnected with the use in eqn. (1)]. α_0 is an arbitrary phase angle specifying the position of the zeros of the pattern with respect to the envelope. Similarly, addition of the forward- and backward-travelling first reverse components yields a second standing wave, and it is evident that this will once more be in step with the fundamental pattern at a point $(m + 1)$ wavelengths further along the line.

Thus the total amplitude of the field from these components will be given by

$$A = a_0 \sin (m\theta + \alpha_0) + a_{-1} \sin [(m + 1)\theta + \alpha_{-1}] \quad . \quad (4)$$

where α_{-1} has the same significance for the first reverse component as does α_0 for the fundamental.

Eqn. (4) can be written

$$A = R\sin(m\theta + \alpha_0 + \phi) \quad . \quad . \quad . \quad (5)$$

where
$$R^2 = a_0^2 + 2a_0a_{-1}\cos\theta' + a_1^2 \quad . \quad . \quad . \quad (6)$$

and
$$\phi = \arc\tan\left(\frac{\sin\theta'}{\dfrac{a_0}{a_{-1}} + \cos\theta'}\right) \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where
$$\theta' = \theta + \alpha_{-1} - \alpha_0 \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The ratio of the maximum to the minimum of the envelope, r say, is given by

$$r = \frac{a_0 + a_{-1}}{a_0 - a_{-1}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

and this equation can be used with the experimental curves to determine the ratio of the amplitude of the fundamental to the first reverse component, as in Fig. 5 (solid line). The zeros of the pattern will be given by

$$s\pi = m\theta' + (m+1)\alpha_0 - m\alpha_{-1} + \phi \quad . \quad . \quad (10)$$

This analysis is formally identical with that for single-sideband modulation theory,[6] and the function giving the variation of the spacing of the zeros of the pattern is sketched in Fig. 6.

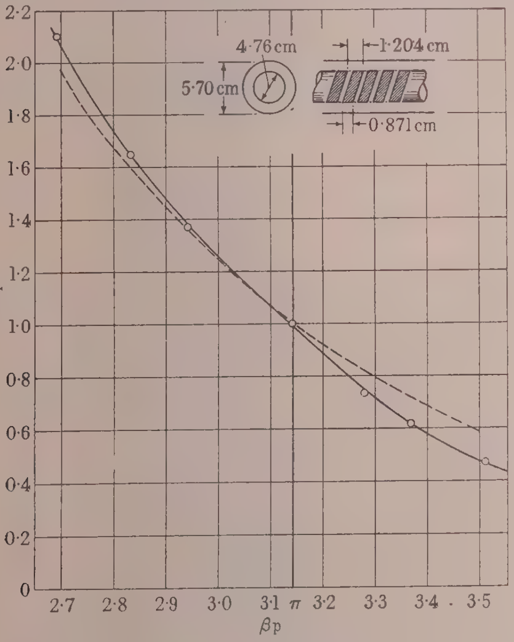


Fig. 5.—Experimental plot of a_0/a_{-1} together with assumed approximation function.

⊙ Experimental points.
– – – Approximation function.

If the argument in eqn. (5), denoted by h, is plotted against θ', a curve will be obtained as shown in Fig. 7. Insertion of the abcissae at ($h = s\pi$) according to eqn. (10) to intersect this curve defines the ordinates for the zeros which exhibit the periodic variations of apparent wavelength. It is seen that at $\theta' = \pi$, i.e. near the minimum of the envelope, the apparent wavelength is increased, while it is decreased near the maximum ($\theta' = 0$ or 2π).

For a correct evaluation of m and thus the fundamental wavelength, the position of at least two zeros is required, together with a knowledge of the amplitude ratio, r, of the envelope, but a more satisfactory procedure is to measure all zero positions and average over complete periods or half periods of the envelope. This was done to obtain the experimental phase diagram of Fig. 8, up to approximately 0·95 of the π-frequency. Since the space-harmonic and fundamental wavelengths are very nearly equal at this point, the envelope is very long (m large), and exceeds the length of travel of the probe. The h, θ' curve of Fig. 7 would thus have a large average gradient, and the addition of ϕ increases this by a small percentage only, up to $\theta' = \pi$, and decreases it slightly from $\theta' = \pi$ to 2π. Over the majority of the envelope length the spacing of the zeros is constant, and in plotting the points of Fig. 8 from 0·95 to 1·05 of the π-frequency it was assumed that this spacing gave the true fundamental wavelength and the correction due to ϕ could be neglected. This gives a phase-change per turn which is very nearly constant at π over a finite band, and thus apparently a dispersive region. It will be shown that the neglect of the discontinuity of spacing of the zeros near the minimum of the envelope is not permissible and accounts for this feature.

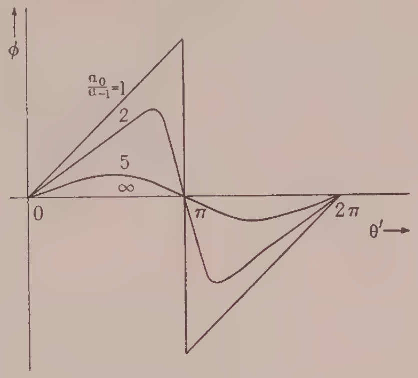


Fig. 6.—Phase-correction function.

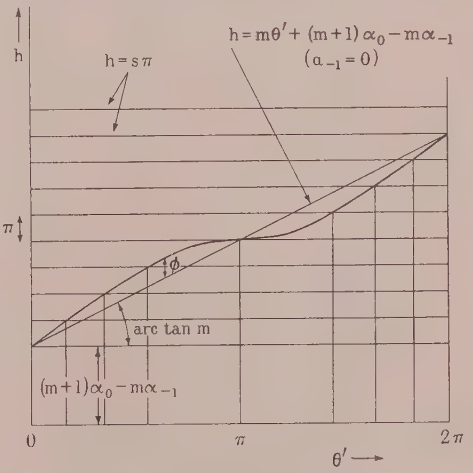


Fig. 7.—Phase-distribution function.

The ratio of the apparent wavelength, λ_m, measured near the maximum of the envelope, to the true wavelength will be given very closely by

$$\frac{\lambda_m}{\lambda} = m\left[\frac{d\theta'}{dh}\right]_{\theta'=0} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

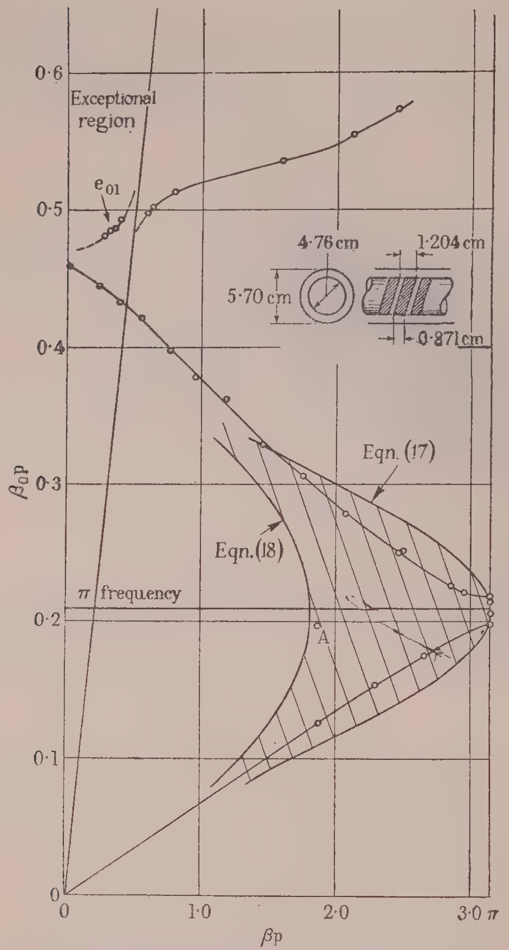


Fig. 8.—Experimental phase diagram for line (shown inset) with uncertainty region (shaded).

In terms of the equivalent phase constants (β_m, β), this becomes

$$\frac{\beta_m}{\beta} = \frac{1}{m}\left[\frac{dh}{d\theta'}\right]_{\theta'=0} \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Using eqn. (7) and the fact that $h = m\theta' + (m+1)\alpha_0 - m\alpha_{-1} + \phi$

$$\frac{\beta m}{\beta} = 1 + \frac{1}{m(a_0/a_{-1}) + 1} \quad . \quad . \quad . \quad . \quad (13)$$

This relation can be used in conjunction with β for an idealized helix characteristic to predict how β_m would depart from the true value. Thus taking

$$\beta = \beta_0 \operatorname{cosec} \psi \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

it is readily shown that

$$m = \frac{\beta_0 p \operatorname{cosec} \psi}{2(\pi - \beta_0 p \operatorname{cosec} \psi} \quad . \quad . \quad . \quad . \quad (15)$$

from which the property that m, and thus the length of the envelope, is infinite at $\beta_0 p \operatorname{cosec} \psi = \pi$ can be seen. To obtain β_m as a function of β_0 from eqn. (13), a knowledge of a_0/a_{-1} is required in addition to eqn. (15). At low frequencies there is no deviation of β_m from β because all space-harmonic amplitudes tend to zero compared with that of the fundamental. For this to be so it follows from eqn. (13) that a_0/a_{-1} must tend to infinity more rapidly than $1/m$ at low frequencies. Here m is proportional to frequency [see eqn. (15)]. The correct dependence of a_0/a_{-1} requires a field theory as developed by Sensiper,[1] extended to cover the case for an outer shield and for the helix wound on a dielectric, as used in the experimental work here.

For a simple treatment a function will be chosen which has a sufficiently high-order pole at zero frequency, and passes through the point (1, π) [Fig. 4B]. It is found that

$$\frac{a_0}{a_{-1}} = \cot^3 \frac{(\beta_0 p, \operatorname{cosec} \psi)}{4} \quad . \quad . \quad . \quad . \quad (16)$$

compares well with experimental results for the particular line which will be used to illustrate the phenomenon, up to the π-frequency, although agreement is not so good above this (see Fig. 5). Using eqns. (13) and (15) with eqn. (16) gives

$$\beta_m p = B\left[1 + \frac{2\left(\frac{\pi}{B} - 1\right)}{\cot^3 \frac{B}{4} + 1}\right] \quad . \quad . \quad . \quad . \quad (17)$$

where $B = \beta_0 p \operatorname{cosec} \psi$.

This relation is shown in Fig. 8. The near-stationary phase-change at $\beta p = \pi$ is in evidence, and it can be concluded that the experimental results are due to this feature and not to dispersion of the helix. A similar procedure can be carried out to predict the experimental curve which would arise if wavelengths were measured only near the minimum of the envelope. This gives

$$\beta'_m p = B\left[1 + \frac{2\left(\frac{\pi}{B} - 1\right)}{\cot^3 B/4 - 1}\right] \quad . \quad . \quad . \quad (18)$$

which is also shown in Fig. 8.

The $\beta_m p$ and $\beta'_m p$ curves enclose a region into which experimental points would lie if the wavelength were measured at random with respect to the envelope of the pattern. It was, in fact, the observation of a random distribution of points over this area when measurements were first carried out on the transmission characteristics of a tape helix in this region which led to an analysis of the phenomena involved.

On the curve of Fig. 5 an experimental point has been inserted for the π-frequency. This can always be obtained by changing the frequency until the envelope of the maxima is constant as far as can be ascertained throughout the length of the line. If the frequency is *exactly* that for phase-change of π per turn, it is possible to rotate the helix about its axis to give a measured pattern which is zero throughout the length of the line. The appearance of this condition is a very sensitive indication that the frequency is accurately that for a phase-change of π per turn.

The curve shown in Fig. 9(*a*) was measured with the last turn of the helix open-circuited 180° from the line of travel of the probe, and so there is a minimum of the envelope at the end of the line. Rotation of the line about its axis by 180°, so as to bring the open-circuit beneath the travelling probe, gives a maximum of the envelope at the end of the line [Fig. 9(*b*)]. As the frequency is not sufficiently close to the π-frequency to extend the envelope beyond the length of available travel, a minimum of the envelope is observable. When the frequency reaches the π-condition this minimum falls to zero and broadens to fill the length of the line along a single azimuthal angle, as indicated in the preceding paragraph.

A point A is shown in Fig. 8 which is obtained from the single wide zero spacing at the minimum of the envelope in Fig. 9(*b*). It would not be expected that this would lie exactly on the predicted boundary, owing to the finite difference approximation implicit in eqn. (12).

To summarize, it is necessary to average wavelengths measured from the standing-wave pattern over integral numbers of half periods of the envelope in order to obtain the correct phase characteristic, but this becomes impossible at some frequency just below that giving a phase change of π per turn, for a finite length of measuring line.

Undoubtedly the simplest procedure is to extrapolate through the π region, but if further detail is required it is necessary to measure the ratio $a_0 : a_{-1}$ and extrapolate this to obtain the true wavelength as described.

(3) EFFECTIVE WAVE VELOCITIES IN A SHORT HELIX

Although the previous Section has been concerned with measurements on a reflectively terminated line, the same features would be observed with measurements on a matched helix using a phase-sensitive detector. Furthermore, a travelling electron stream will itself be subject to disturbances from the same phenomenon. This may be regarded as being due to the finite length of the helix, within which the effect of space harmonics other than the desired component does not average to zero.

For a travelling-wave oscillograph, for which the measurements described were undertaken, it is laborious but not difficult

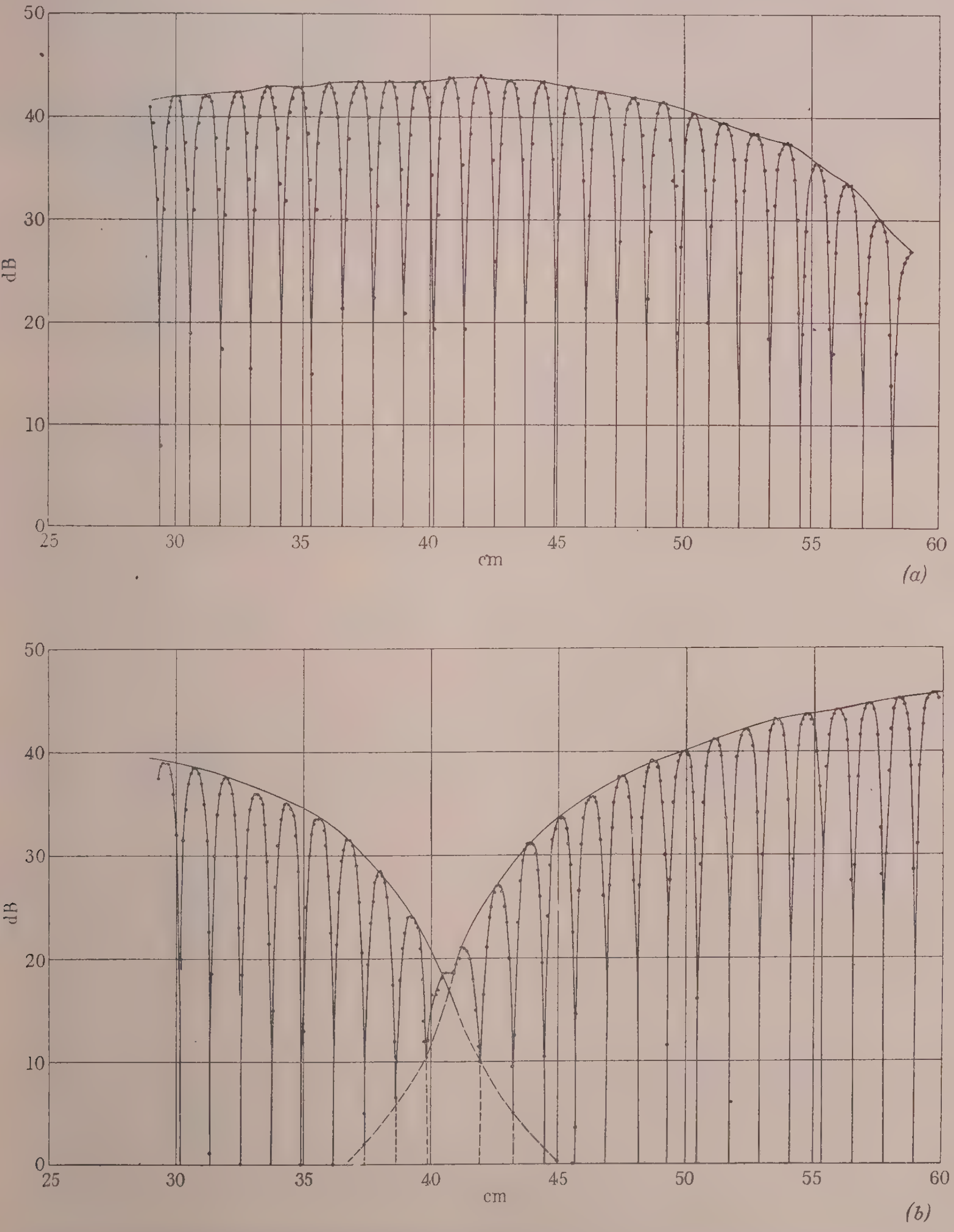


Fig. 9.—Experimental standing-wave patterns taken near the π frequency for the line used in Fig. 8.

to calculate the deflection characteristic for a short helix, assuming small sinusoidal r.f. fields, and taking into account the fundamental and first reverse component. However, this is of doubtful value, since such instruments are, in general, required for the observation of non-sinusoidal waveforms, for which the indicated calculation gives only indirect information.

It is considered more likely that the limit to the high-frequency performance of a helical-line deflection system will be set more by the decrease in amplitude of the fundamental component with frequency as more of the energy is propagated in the higher space-harmonic fields.

(4) PROPAGATION IN THE EXCEPTIONAL REGION

As mentioned in Section 1, measurements were extended sufficiently high in frequency to cross the $m = -1$ exceptional region. When measuring with the method used, the longest wavelength component is obtained which from $\beta p = \pi$ to 2π is the first reverse component. The corresponding fundamental branch is obtained by taking the image of these points in the $\beta p = \pi$ ordinate, but this has not been done in Fig. 8. It is seen that points are obtained which agree with the e_{01} mode predicted by Stark, but the cut-off before the boundary of the exceptional region is not present.

(5) REFERENCES

(1) Sensiper, S.: "Electromagnetic Wave Propagation on Helical Structures. A Review and Survey of Recent Progress," *Proceedings of the Institute of Radio Engineers*, 1955, **43**, p. 149.

(2) Stark, L.: "Lower Modes of a Concentric Line having a Helical Inner Conductor," *Journal of Applied Physics*, 1954, **25**, p. 1155.

(3) Lines, A. W., *et al.*: "Some Properties of Waveguides with Periodic Structures," *Proceedings I.E.E.*, Paper No. 941 R, July, 1950 (**97**, Part III, p. 263).

(4) British Patent No. 722217.

(5) Epzstein, B., and Mourier, G.: Definition, mesure et caractères des vitesses de phase dans les systèmes a structure periodique," *Annales de Radioélectricité*, 1955, **10**, p. 64.

(6) Reich, H. J.: "Theory and Application of Electron Tubes" (McGraw-Hill, New York, 1944), p. 315.

21.396.11:538.566

Monograph No. 152 R
Oct. 1955

AN APPROXIMATE THEORY OF THE DIFFRACTION OF AN ELECTROMAGNETIC WAVE BY AN APERTURE IN A PLANE SCREEN

By R. F. MILLAR, M.A.

(*The paper was first received* 18*th May, and in revised form* 19*th July*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in October*, 1955.)

SUMMARY

An approximate diffraction theory based on the Sommerfeld half-lane solution is developed. It is shown that, in certain regions, the lectromagnetic field scattered (in the two-dimensional problem) by a erfectly conducting half-plane with plane waves incident can be onceived as arising from the flow of electric and magnetic currents long the edge of the half-plane. This *edge-current* concept is extended o the diffraction of normally incident plane waves by an aperture of rbitrary form in a thin, perfectly conducting screen of infinite extent.

The approximation is for large apertures and is probably asymptotic. nteraction between the fields scattered by different parts of the screen s neglected in the present treatment.

Expressions are obtained for the tangential electric field in the aperure in terms of these edge currents. The case of the circular aperture s studied in some detail, axial and aperture field distributions being erived and compared with the available experimental and theoretical ata. Agreement with experiment and theory is fairly close, and mproves with increase of the ratio of aperture radius to wavelength of ncident radiation.

LIST OF SYMBOLS

(x, y, z) = Rectangular Cartesian co-ordinates of point of observation.
λ = Wavelength of incident radiation.
ω = Angular frequency of incident radiation.
c = Velocity of light in free space.
$k = \omega/c$.
$E^i(z)$ = Incident electric field of unit intensity.
$E(x, y, z)$ = Total electric field.
$E_x(x, y, z), E_y(x, y, z), E_z(x, y, z)$ = Components of $E(x, y, z)$.
$H^i(z)$ = Incident magnetic field.
$H(x, y, z)$ = Total magnetic field.
$H_x(x, y, z), H_y(x, y, z), H_z(x, y, z)$ = Components of $H(x, y, z)$.
ϵ_0 = Permittivity of free space.
μ_0 = Permeability of free space.
$Z = \sqrt{(\mu_0/\epsilon_0)}$.
P = Point of observation, with co-ordinates (x, y, z).
T = Projection of P on $z = 0$.
Q = Point on rim of aperture.
R = Distance between P and Q.
ρ = Distance between T and Q.
ds = Element of rim of aperture at Q.
ds = Magnitude of ds.
γ = Angle between positive x-axis and TQ.
θ = Angle between $E^i(0)$ and ds.
ψ = Angle between aperture plane and plane containing ds, PQ.
(r, ϕ) = Polar co-ordinates of point of observation in half-plane discussion.
χ = Angle between $E^i(0)$ and edge of half-plane.
E_1, E_2 = Electric fields scattered by half-plane when $\chi = 0$, $\frac{1}{2}\pi$, respectively.
I_1^e, I_1^m = Electric and magnetic edge currents, respectively.
E_1^e, E_1^m = Electric far fields produced by I_1^e, I_1^m, respectively.
I^e, I^m = Edge currents corresponding to arbitrary χ.
δ = Error introduced by considering field scattered by half-plane as an edge wave.
a = Radius of circular aperture.
$\Pi(x, y, 0)$ = Hertz vector of electric edge current, in circular aperture.
$\Pi_x(x, y, 0), \Pi_y(x, y, 0)$ = Components of $\Pi(x, y, 0)$ in circular aperture.
(r_p, θ_p) = Polar co-ordinates of point in circular aperture.
I_1, I_2, I_3 = Integrals involved in calculation of Π_x, Π_y.
$\beta = \theta - \theta_p$.
$J_n(x)$ = Bessel function of order n and argument x of the first kind.

(1) INTRODUCTION

The development of microwave techniques in recent years stimulated the study, both experimental and theoretical, of many problems in electromagnetic diffraction. Recent measurements at microwave frequencies[1,2,3,4] have indicated the extent to which the classical theory (i.e. Kirchhoff and its electromagnetic extensions) is acceptable, and, in particular, have shown that it yields an exceedingly poor approximation to the fields in the neighbourhood of the diffracting aperture.

The rigorous solution of Maxwell's equations subject to prescribed boundary values has been obtained for relatively few diffraction problems. In particular, the only such solution involving a finite aperture in a plane screen is that given by Meixner and Andrejewski[5,6] for the diffraction of a normally incident plane wave by a circular aperture in a thin, perfectly conducting, plane screen of infinite extent. The method (expansion of the Hertz vector in spheroidal wave functions) lends itself to computations for apertures which are small compared to the wavelength. For large apertures, however, convergence is slower and the computational difficulties are formidable.

A simple means of estimating the field components, especially in the neighbourhood of the aperture, would therefore seem to be of value. Most approximate theories that have been developed are useful only when the maximum dimension of the (bounded) diffracting aperture is less than, or comparable to, the wavelength of the incident radiation. Exceptions are the scalar theory of Braunbek[7,8,9] (extended to electromagnetic diffraction by Frahn[10]), and the geometrical-optics approximation of Bekefi,[11] which may be applied when the smallest dimension of the aperture is large compared to the wavelength. Braunbek's idea was to put the unknown boundary values equal to suitably modified quantities obtained from the rigorous Sommerfeld half-plane solution. Bekefi solved a boundary-value problem for a one-component Hertz vector, which is equivalent to the assumption of geometrical-optics currents on the screen. (Geometrical-optics currents are the currents that would flow in the screen in the absence of the aperture.)

The new method is also based upon the Sommerfeld half-plane solution, but in a different way from that of Braunbek. An

Correspondence on Monographs is invited for consideration with a view to ublication.

Mr. Millar is at the Cavendish Laboratory, Cambridge. He is on leave of absence om the Division of Radio and Electrical Engineering, National Research Council of anada.

examination of the two-dimensional solution to the perfectly conducting half-plane problem shows that, for incident plane waves, the field in part of the illuminated (in the sense of geometrical optics) region appears to radiate from the edge of the half-plane, and thus to be produced by currents (electric and magnetic) flowing along the edge. This edge-current concept is extended to the case of diffraction of normally incident plane waves by a large, arbitrarily shaped aperture in a thin, perfectly conducting screen. Expressions are obtained for the field quantities in part of the illuminated region in terms of relatively simple line integrals around the edge of the aperture. Although similar (in that they are line integrals) to the empirical formulae

$$E_x = -\frac{1}{2\pi}\oint \frac{\varepsilon^{-jk\rho}}{\rho}\cos\theta\sin\gamma ds$$

$$E_y = 1 - \frac{1}{2\pi}\oint \frac{\varepsilon^{-jk\rho}}{\rho}\cos\theta\cos\gamma ds$$

developed by Andrews[1] to describe his experimental results, they are significantly different.

It may be worth while to emphasize the hypothetical nature of these edge currents, and to distinguish them from the current filaments introduced by Moullin and Phillips[12] to replace the excess current near the edge of a half-plane or ribbon. These authors found that the effect of such an edge on the current density is negligible, except at points relatively close to it where the density becomes singular. The effect of this excess current could be accounted for by the use of these current filaments situated on or near the edge.

On the contrary, the present theory makes no claim to approximate to the currents near the edge in this way. The edge-current concept and terminology are introduced only because of the form of the leading term in the asymptotic expansion of the half-plane field, and the strength of the edge currents will be found to depend on the point from which they are observed.

Since the derivation of the edge currents is based on the asymptotic expansion of the half-plane field, it is considered likely that the approximation, where valid, is asymptotic to the true field. Evaluation of eqns. (8) for the aperture distribution by the method of steepest descents yields, in the limiting case of a large circular aperture, the leading term in the asymptotic form of the half-plane field.

The effect of interaction between the fields scattered by different parts of the screen is neglected in the following treatment. The method is used to determine an approximation to the field on the axis of a circular aperture and in the aperture itself. These results are compared with experiment and with the available theoretical data.

(2) DEFINITIONS AND NOTATION

In the subsequent analysis, it will be assumed that the diffracting aperture and plane screen occupy the plane $z = 0$ in an orthogonal system of Cartesian co-ordinates. The incident plane wave falls normally on to the screen and aperture from the negative z-direction with z-dependence $\exp(-jkz)$. The propagation coefficient, k, is equal to $2\pi/\lambda$, where λ is the wavelength of the incident radiation. M.K.S. rationalized units are used, and time dependence $\exp(j\omega t)$ is understood for all field quantities. Impedance is defined in terms of the permittivity ϵ_0 and permeability μ_0 of free space by $Z = \sqrt{(\mu_0/\epsilon_0)}$. The incident and total electric fields are denoted by $E^i(z) = E^i(0)\exp(-jkz)$ and $E(x, y, z)$ respectively, with corresponding notation for the magnetic fields: $H^i(z) = H^i(0)\exp(-jkz)$ and $H(x, y, z)$.

In Fig. 1 are illustrated most of the symbols used in the discussion of diffraction by an aperture of arbitrary form. P is the point of observation and T its orthogonal projection onto the aperture plane. R is the distance from P to a point Q on the rim of the aperture, whilst ρ is the distance from T to Q. γ is the angle between the positive x-axis and TQ, and ds is the element of edge at Q, oriented in the direction of increasing γ. θ is the angle between $E^i(0)$ and ds, and ψ is the angle (greater than or equal to π for points on the "shadow side," i.e. $z \geqslant 0$) between the aperture plane and the plane containing ds, PQ.

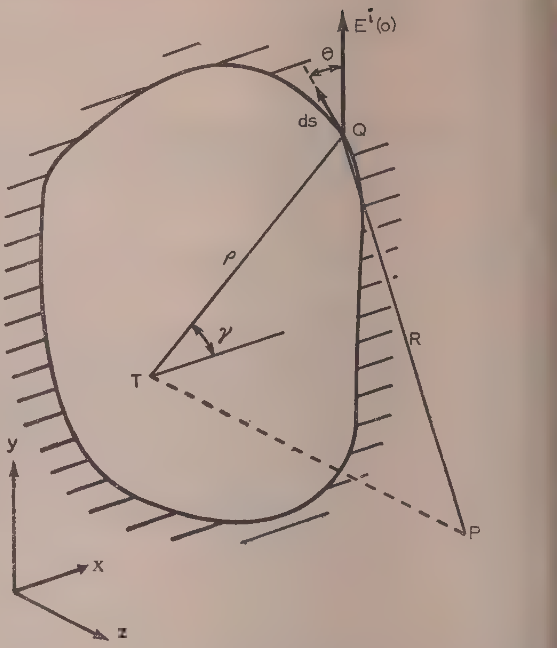


Fig. 1.—Aperture of arbitrary form in a plane screen, with associated symbols.

In the discussion of the Sommerfeld half-plane solution, r is the perpendicular distance from a point P to the edge of the half-plane occupying $z = 0$, $x \geqslant 0$. ϕ is the polar angle of P relative to the positive x-axis.

(3) DERIVATION OF EDGE CURRENTS

(3.1) Perfectly conducting Half-Plane

The Sommerfeld solution to the two-dimensional problem of the diffraction of a normally incident plane wave by a thin, perfectly conducting half-plane ($z = 0$, $x \geqslant 0$) is first examined.

The electric field at a point P ($z \geqslant 0$, $x < 0$) with polar co-ordinates (r, ϕ) scattered by the half-plane is considered for two particular cases of normal incidence. A suitable linear combination of these two will then give the scattered field for arbitrary incident polarization.

Case 1.—Incident electric vector parallel to the y-axis, or E-polarization.

It is well known (e.g. Reference 13, pp. 145–148) that the first term of the asymptotic expansion for the scattered electric field at P is

$$E_1 = (0, 1, 0)\frac{\varepsilon^{-j(\pi/4+kr)}}{\sqrt{(\pi kr)}}\sec\phi\sin\tfrac{1}{2}\phi \quad . \quad . \quad . \quad (1)$$

with an error less than $\frac{1}{2}\pi\delta^3$ if P lies outside the parabola $2\pi\delta^2 kr\cos^2\frac{1}{2}(\phi - \frac{1}{2}\pi) = 1$.

For small error this condition implies that $kr\cos^2\frac{1}{2}(\phi - \frac{1}{2}\pi) \gg 1$. Therefore, as well as being in the far field ($kr \gg 1$), P must not be too near the edge of the geometrical shadow ($\phi = \frac{3}{2}\pi$).

Case 2.—Incident magnetic vector parallel to the *y*-axis, or *l*-polarization.

In this case the incident electric vector is parallel to the -axis, and the scattered electric field at P is

$$E_2 = (\sin\phi, 0, \cos\phi)\frac{\varepsilon^{-j(\pi/4+kr)}}{\sqrt{(\pi kr)}}\sec\phi\cos\tfrac{1}{2}\phi \quad . \quad . \quad (2)$$

ith the error again less than $\frac{1}{3}\pi\delta^3$ for P situated as before.

The far fields radiated by an electric current I_1^e and a magnetic urrent I_1^m, both independent of y and flowing along the y-axis, re as follows:

$$\left.\begin{aligned} E_1^e &= (0, 1, 0)\frac{kZ}{2\sqrt{2}}I_1^e\frac{\varepsilon^{-j(3\pi/4+kr)}}{\sqrt{(\pi kr)}} \\ \text{nd} \qquad E_1^m &= (\sin\phi, 0, \cos\phi)\frac{k}{2\sqrt{2}}I_1^m\frac{\varepsilon^{j(\pi/4-kr)}}{\sqrt{(\pi kr)}} \end{aligned}\right\} \quad . \quad . \quad (3)$$

Thus the fields E_1 and E_2 at (r, ϕ) with $\pi \leqslant \phi < \frac{3}{2}\pi$, are ust those that would be produced by electric and magnetic urrents I_1^e and I_1^m respectively, flowing along the y-axis, where

$$\left.\begin{aligned} I_1^e &= \frac{2\sqrt{2}j}{kZ}\sec\phi\sin\tfrac{1}{2}\phi \\ I_1^m &= -\frac{2\sqrt{2}j}{k}\sec\phi\cos\tfrac{1}{2}\phi \end{aligned}\right\} \quad . \quad . \quad . \quad (4)$$

An arbitrarily-polarized normally-incident plane wave may be epresented by $E^i(0) = (\sin\chi, \cos\chi, 0)$, where χ is the angle etween $E^i(0)$ and the positive y-axis. $E^i(z)$ is thus the sum of wo waves: one H-polarized and proportional to $\sin\chi$, and the ther E-polarized and proportional to $\cos\chi$.

By superposition of the previous results, the electric and magetic currents flowing along the y-axis that correspond to this ncident electric field are, respectively,

$$\left.\begin{aligned} I^e &= \frac{2\sqrt{2}j}{kZ}\cos\chi\sec\phi\sin\tfrac{1}{2}\phi \\ I^m &= -\frac{2\sqrt{2}j}{k}\sin\chi\sec\phi\cos\tfrac{1}{2}\phi \end{aligned}\right\} \quad . \quad . \quad . \quad (5)$$

with the same limit to the error in the field as before.

(3.2) Aperture of Arbitrary Form

The foregoing results will now be extended to the case of an perture of arbitrary form in a perfectly conducting screen ccupying the plane $z = 0$. The following seemingly plausible ssumptions are made:

(*a*) In the part of the illuminated zone ($z \geqslant 0$), not too close to the edge of the geometrical optics shadow and not too far from the plane $z = 0$, the field scattered by the screen is mainly an edge wave, which could be produced by the flow of electric and magnetic currents on the rim of the aperture.

(*b*) The hypothetical currents flowing in an element ds of the rim of the aperture [where ds makes an angle χ with $E^i(0)$] are the same as were derived previously for a half-plane lying in the plane of the screen, with edge making an angle χ with $E^i(0)$.

Although the far diffracted field (i.e. the difference between the otal and geometrical-optics far fields) of the half-plane is an dge wave in regions of space other than the illuminated portion f $z \geqslant 0$, discussion in the case of an arbitrary aperture is conined to the latter zone because of difficulties in the extension of he edge-current concept across the shadow boundary.

If P were too far from the plane of the screen, or too near the dge of the aperture, effects not attributable to these simple edge urrents would become increasingly important.

The second assumption, which identifies the aperture edge current in an element ds with that for a correspondingly oriented half-plane, should be qualified somewhat. It will be valid only if the curvature of the edge of the aperture is small. In other words, at each point on the rim, the edge is assumed to be locally straight.

With this assumption, the equivalent edge currents flowing in an element ds (see Fig. 1) of the edge of the aperture to produce the scattered field at P are given by eqns. (5), with $\chi = \theta$, $\phi = \psi$.

$$\left.\begin{aligned} I^e &= \frac{2\sqrt{2}j}{kZ}\cos\theta\sec\psi\sin\tfrac{1}{2}\psi \\ I^m &= -\frac{2\sqrt{2}j}{k}\sin\theta\sec\psi\cos\tfrac{1}{2}\psi \end{aligned}\right\} \quad . \quad . \quad . \quad (6)$$

Each current element produces a field at P, the total scattered field being found by integration over all such elements, or from the Hertz vectors of the currents. (In this latter case, ψ should be considered constant as regards differentiations with respect to the co-ordinates of P, since the factors dependent on ψ should enter only into the final expressions for E and H.)

For a point P in the aperture,

$$\sec\psi\sin\tfrac{1}{2}\psi = -1, \quad \sec\psi\cos\tfrac{1}{2}\psi = 0$$

which implies that, in this approximation, the field is obtained from an electric edge current only. In fact

$$I^e = -\frac{2\sqrt{2}j}{kZ}\cos\theta; \quad I^m = 0 \quad . \quad . \quad . \quad . \quad (7)$$

With $E^i(0) = (0, 1, 0)$, the total far field at P (e.g. Reference 14, p. 440) is then

$$\left.\begin{aligned} E_x(x, y, 0) &= \frac{1}{\sqrt{2\pi}}\oint\frac{\varepsilon^{-jk\rho}}{\rho}\cos(\gamma-\theta)\cos\theta\sin\gamma ds \\ E_y(x, y, 0) &= 1 - \frac{1}{\sqrt{2\pi}}\oint\frac{\varepsilon^{-jk\rho}}{\rho}\cos(\gamma-\theta)\cos\theta\cos\gamma ds \end{aligned}\right\} \quad . \quad (8)$$

The similarity between Andrews's formulae quoted in Section 1 and eqns. (8) is at once evident, but two significant differences are noticed, namely a phase difference of π in E_x and an added factor $\sqrt{2}\cos(\gamma-\theta)$ in the integrands of eqns. (8). Although measurements of $|E_x|$ would not resolve the question of phase difference, the additional factor should produce a measurable difference between predictions of the two theories. This will be discussed in more detail in Section 5.

(4) CIRCULAR APERTURE

For the remainder of the paper the diffraction of normally incident plane waves by a circular aperture of radius a will be considered. The origin of co-ordinates is taken at the centre of the aperture, and the incident electric field is given by $E^i(0) = (0, 1, 0)$. The electric field only is calculated on the axis—calculations for the axial magnetic field being very similar —and in the aperture itself.

(4.1) Axial Field

If the point P is on the axis, the electric far field may be found most easily by integration over all the current elements. The ψ-dependent factors are now independent of the variable of integration as follows:

$$\left.\begin{aligned} \sec\psi\sin\tfrac{1}{2}\psi &= -\frac{R}{\sqrt{2a}}\sqrt{\left(1+\frac{a}{R}\right)} \\ \sec\psi\cos\tfrac{1}{2}\psi &= \frac{R}{\sqrt{2a}}\sqrt{\left(1-\frac{a}{R}\right)}, \quad R = \sqrt{(a^2+z^2)} \end{aligned}\right\} \quad . \quad (9)$$

The differentials of the electric and magnetic current contributions to E_y [see Reference 14, p. 440, eqn. (2)] are

$$-\frac{1}{2\pi}\sqrt{\left(1+\frac{a}{R}\right)}\varepsilon^{-jkR}\cos^2\theta d\theta$$

$$-\frac{1}{2\pi}\sqrt{\left(1-\frac{a}{R}\right)}\frac{z}{R}\varepsilon^{-jkR}\sin^2\theta d\theta$$

respectively. θ, defined previously as the angle between $E^i(0)$ and ds, is also the polar angle of the variable point on the rim with respect to the positive x-axis. Hence

$$E_y(0, 0, z) = \varepsilon^{-jkz} - \frac{\varepsilon^{-jkR}}{2}\left[\sqrt{\left(1+\frac{a}{R}\right)} + \sqrt{\left(1-\frac{a}{R}\right)}\frac{z}{R}\right] \quad . \quad (10)$$

whilst

$$E_x(0, 0, z) = E_z(0, 0, z) = 0 \quad . \quad . \quad . \quad . \quad (11)$$

Similarly

$$ZH_x(0, 0, z) = -\varepsilon^{-jkz} + \frac{\varepsilon^{-jkR}}{2}\left[\sqrt{\left(1-\frac{a}{R}\right)} + \sqrt{\left(1+\frac{a}{R}\right)}\frac{z}{R}\right] \quad . \quad . \quad . \quad . \quad (12)$$

with

$$H_y(0, 0, z) = H_z(0, 0, z) = 0 \quad . \quad . \quad . \quad . \quad (13)$$

It is of interest to examine eqn. (10) in more detail, since the transmission coefficient, t, of the aperture may be simply related to the far axial field (see References 15 and 16). For a circular aperture of radius a in a thin, plane screen, irradiated by normally incident plane waves polarized parallel to the y-axis, the relevant theorem states that

$$t = \lim_{z\to\infty}\frac{2}{a^2}\mathscr{R}\left[\frac{E_y(0, 0, z)}{jk\,\dfrac{\varepsilon^{-jkz}}{z}}\right] \quad . \quad . \quad . \quad . \quad (14)$$

If $|z| > a$, the expression for E_y may be expanded in powers of a, giving

$$E_y(0, 0, z) = \frac{jka^2}{2}\frac{\varepsilon^{-jkz}}{z} + 0\left(\frac{1}{z^2}\right) \quad . \quad . \quad . \quad (15)$$

On inserting eqn. (15) for $E_y(0, 0, z)$ into eqn. (14), the value $t = 1$ is obtained for all values of ka. Little significance can be attached to this, however, because for large values of z/a the point of observation approaches the border of the geometrical-optics shadow, the first assumption of Section 3.2 is violated, and the approximation method is suspect.

To obtain an improvement over the geometrical-optics value of unity it seems likely that the previously mentioned interaction must be taken into account, and an approximation valid for large values of z/a must be used.

(4.2) Tangential Electric Field in the Aperture

Approximate expressions will now be obtained for the (tangential) components of the electric field in the aperture at a point not too close to the rim. As noted earlier, only the electric edge current contributes to the scattered aperture field. This will be calculated from the Hertz vector, Π, of the edge current in order to illustrate the alternative approach, and because the resulting integrals were found more suitable than eqns. (8) for numerical integration.

The electric Hertz vector for the scattered field has components

$$\left.\begin{aligned}\Pi_x(x, y, 0) &= \frac{a}{2\sqrt{2\pi k^2}}(I_1 - I_2)\sin 2\theta_p\\ \Pi_y(x, y, 0) &= -\frac{a}{\sqrt{2\pi k^2}}(I_1\cos^2\theta_p + I_2\sin^2\theta_p)\end{aligned}\right\} \quad . \quad ,(16)$$

and

$$\operatorname{div}\Pi(x, y, 0) = \frac{I_3}{\sqrt{2\pi k^2}}\sin\theta_p \quad . \quad . \quad . \quad (17)$$

(see Section 8)

where

$$\left.\begin{aligned}I_1 &= \int_0^{2\pi}\frac{\varepsilon^{-jk\rho}}{\rho}\cos^2\beta d\beta\\ I_2 &= \int_0^{2\pi}\frac{\varepsilon^{-jk\rho}}{\rho}\sin^2\beta d\beta\\ I_3 &= \int_0^{2\pi}\frac{\varepsilon^{-jk\rho}}{\rho}\cos\beta d\beta\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (18)$$

$$\rho^2 = r_p^2 + a^2 - 2r_p a\cos\beta,\ \beta \doteq \theta - \theta_p \quad . \quad . \quad (19)$$

(r_p, θ_p) are polar co-ordinates of the point P in the aperture.

Exact evaluation of the integrals I_1, I_2 and I_3 in infinite series is possible by methods similar to those employed by Bekefi[11] and Stenzel[17]. However, to obtain the greatest accuracy with the fewest terms, approximate methods are used.

The integrals may be evaluated approximately for $r_p/a \ll 1$ in terms of Bessel functions. Different approximations to ρ and $1/\rho$ are made, depending on the values of β for which the moduli of the integrands are greatest. For I_1 and I_3, the approximations $\rho \simeq a(1 - r/a\cos\beta)$ and $1/\rho \simeq (1 + r/a\cos\beta)/a$ are appropriate. Similarly $\rho \simeq a[\alpha - r/(a\alpha)\cos\beta]$, $1/\rho \simeq [1 + r/(a\alpha)^2\cos\beta]/(a\alpha)$, with $\alpha^2 = 1 + (r/a)^2$, is chosen for I_2.

With the above approximations and use of the relation

$$\mathrm{J}_n(x) = \frac{j^{-n}}{2\pi}\int_0^{2\pi}\varepsilon^{jx\cos\beta + jn\beta}d\beta \quad . \quad . \quad . \quad . \quad (20)$$

the following results are obtained:

$$\left.\begin{aligned}I_1 &\simeq \pi\frac{\varepsilon^{-jka}}{a}\left\{\mathrm{J}_0(kr_p) - \mathrm{J}_2(kr_p) + \frac{j}{2}\frac{r_p}{a}[3\mathrm{J}_1(kr_p) - \mathrm{J}_3(kr_p)]\right\}\\ I_2 &\simeq \pi\frac{\varepsilon^{-jka\alpha}}{a\alpha}\left[\mathrm{J}_0(kr_p/\alpha) + \mathrm{J}_2(kr_p/\alpha) + \frac{4j}{ka\alpha^2}\mathrm{J}_2(kr_p/\alpha)\right]\\ I_3 &\simeq 2\pi j\frac{\varepsilon^{-jka}}{a}\left\{\mathrm{J}_1(kr_p) + \frac{j}{4}\frac{r_p}{a}[\mathrm{J}_0(kr_p) - \mathrm{J}_2(kr_p)]\right\}\end{aligned}\right\}(21)$$

Hence the total tangential electric field in the aperture, calculated on the edge-current hypothesis, is as follows

$$E_x(x, y, 0) = \frac{\sin 2\theta_p}{2\sqrt{2}}\left[\varepsilon^{-jka}\left\{\mathrm{J}_0(kr_p) - \mathrm{J}_2(kr_p) + \frac{j}{2}\frac{r_p}{a}[3\mathrm{J}_1(kr_p) - \mathrm{J}_3(kr_p)]\right\} - \frac{\varepsilon^{-jka\alpha}}{\alpha}[\mathrm{J}_0(kr_p/\alpha) + \mathrm{J}_2(kr_p/\alpha)]\right] + 0\left(\frac{1}{ka}\right) \quad . \quad . \quad . \quad (22)$$

$$E_y(x, y, 0) = 1 - \frac{1}{\sqrt{2}}\left[\varepsilon^{-jka}\cos^2\theta_p\left\{\mathrm{J}_0(kr_p) - \mathrm{J}_2(kr_p) + \frac{j}{2}\frac{r_p}{a}[3\mathrm{J}_1(kr_p) - \mathrm{J}_3(kr_p)]\right\} + \frac{\varepsilon^{-jka\alpha}}{\alpha}\sin^2\theta_p[\mathrm{J}_0(kr_p/\alpha) + \mathrm{J}_2(kr_p/\alpha)]\right] + 0\left(\frac{1}{ka}\right) . \quad (23)$$

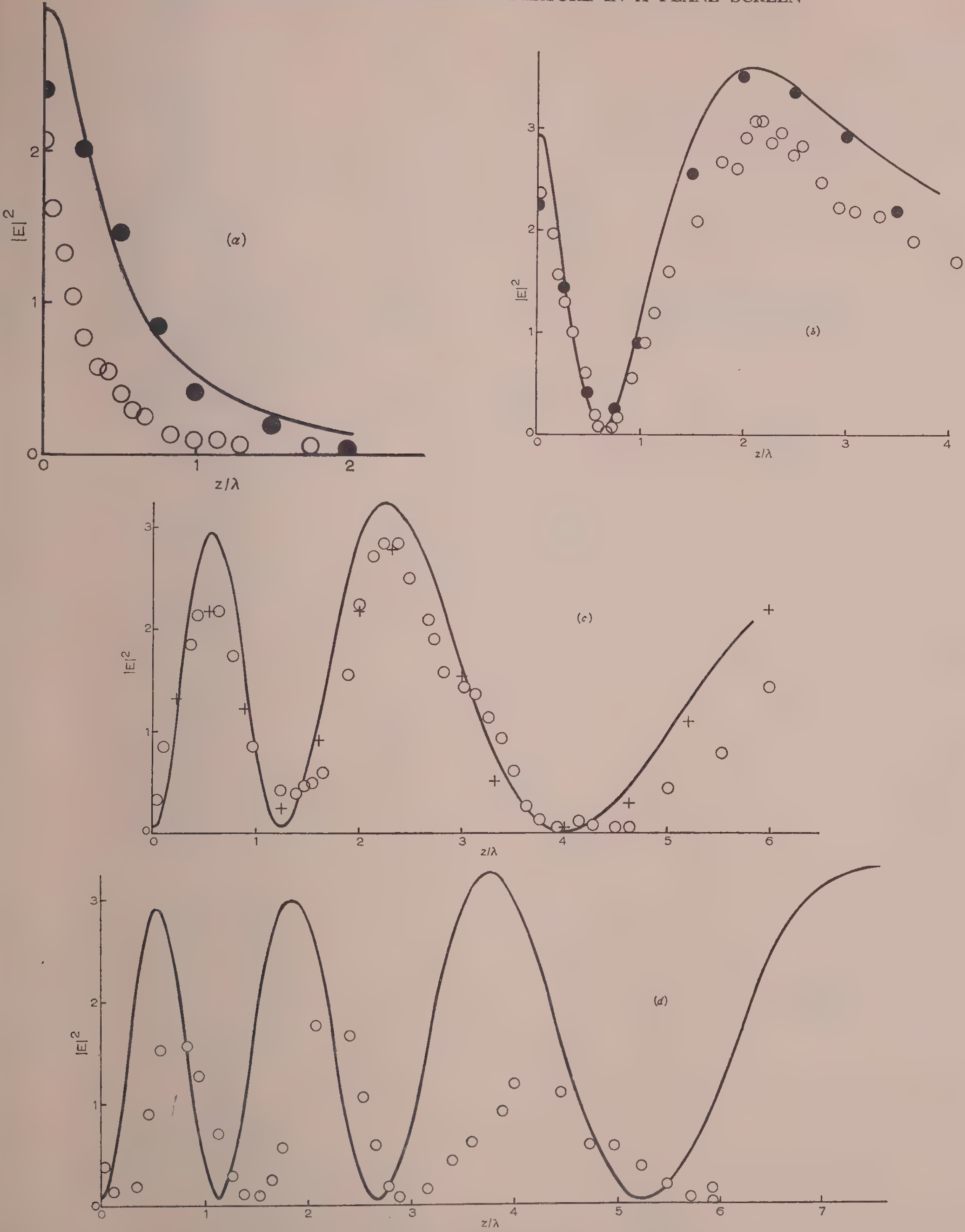


Fig. 2.—Axial intensity of the electric field for circular apertures.

——— Edge-current theory.
● ● Experimental points of Andrews.
○ ○ Experimental points of Silver *et al.*
+ + Experimental points of Severin.

(*a*) Radius $\frac{1}{2}\lambda$.
(*b*) Radius $\frac{3}{2}\lambda$.
(*c*) Radius 3λ.
(*d*) Radius 5λ.

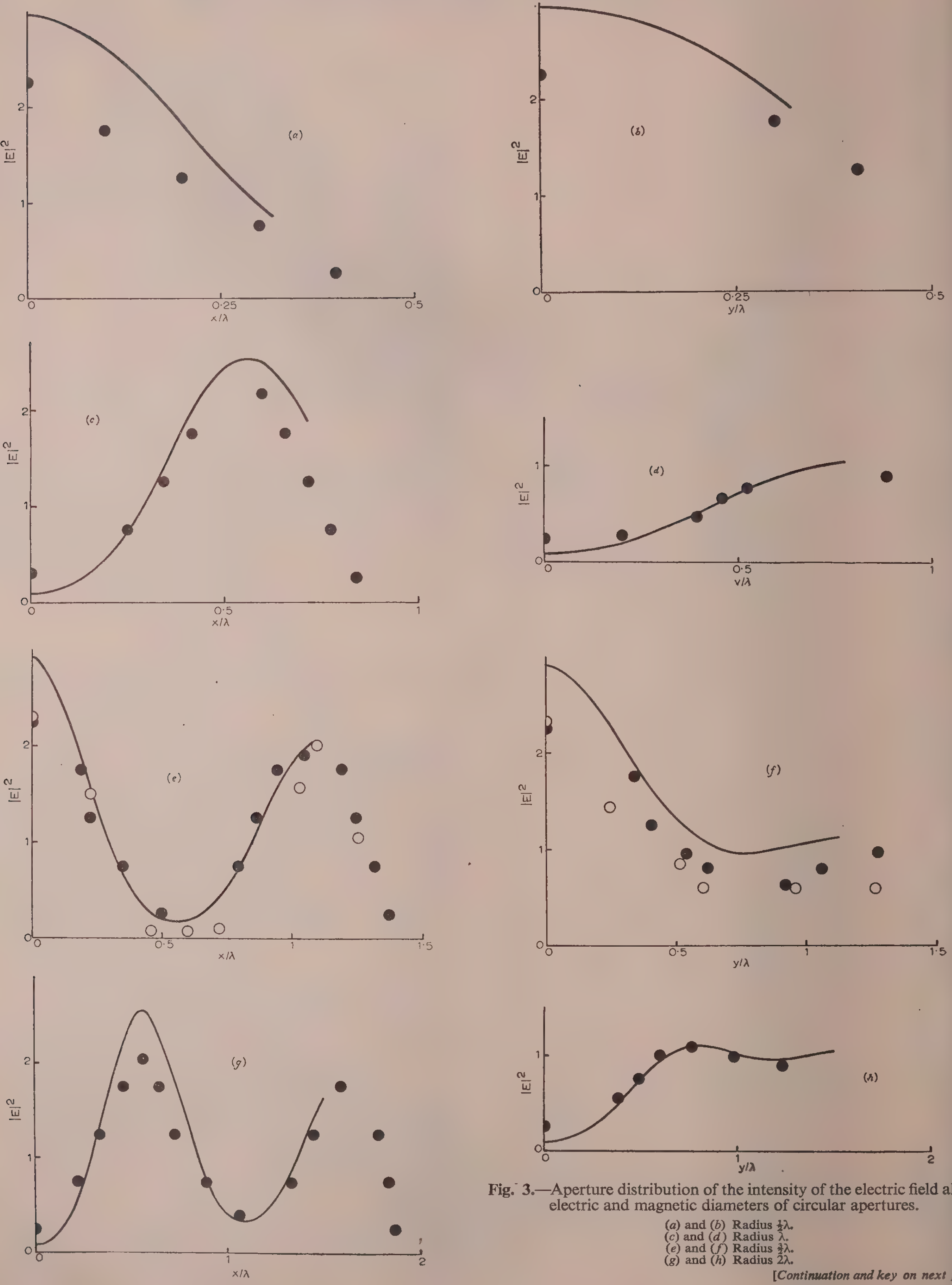


Fig. 3.—Aperture distribution of the intensity of the electric field along electric and magnetic diameters of circular apertures.

(*a*) and (*b*) Radius $\frac{1}{2}\lambda$.
(*c*) and (*d*) Radius λ.
(*e*) and (*f*) Radius $\frac{3}{2}\lambda$.
(*g*) and (*h*) Radius 2λ.

[*Continuation and key on next page.*

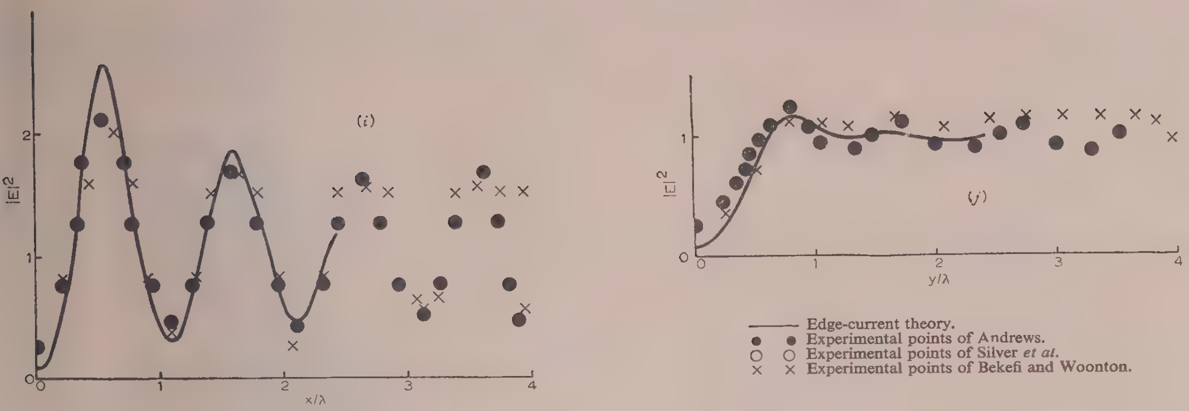


Fig. 3—(*continued*)
(*i*) and (*j*) Radius 4λ.

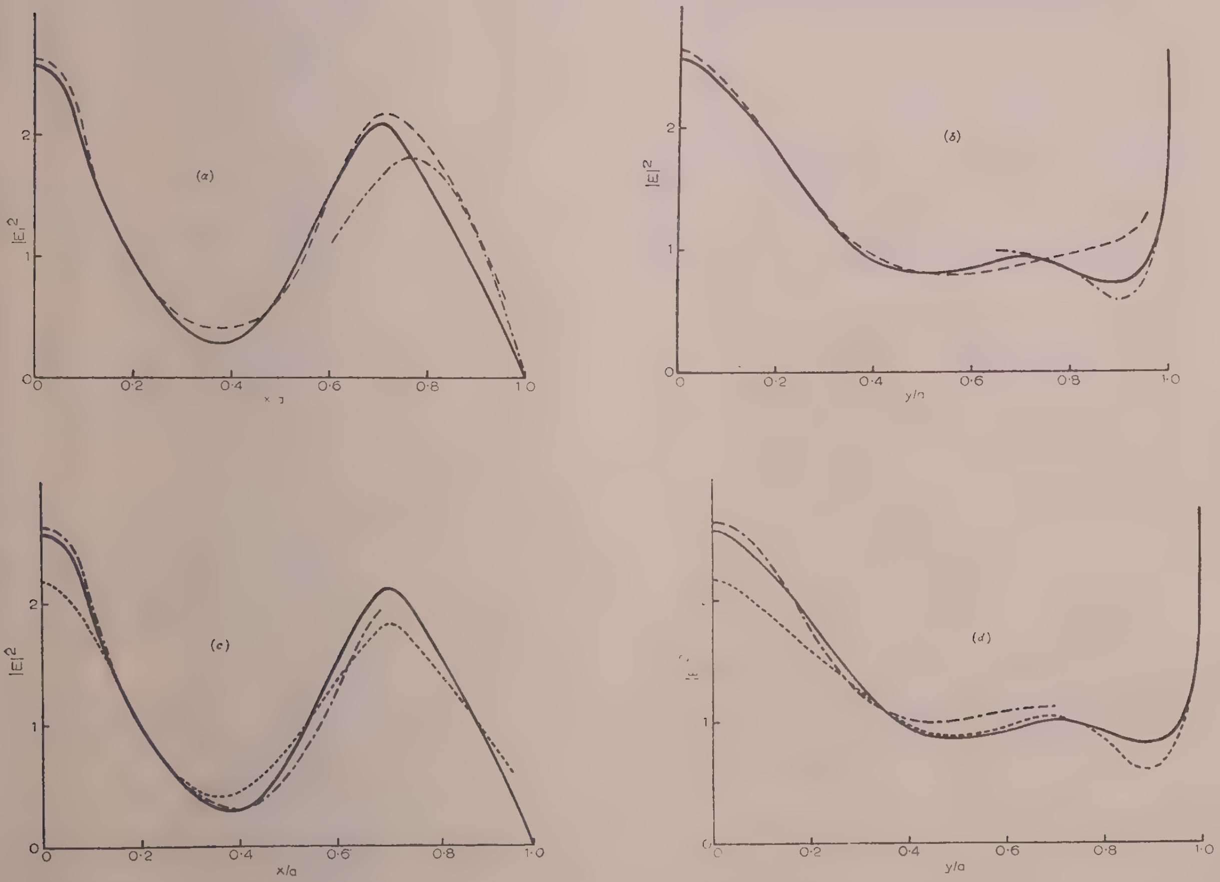


Fig. 4.—Comparison between exact, edge-current, half-plane and geometrical optics-currents theories for the aperture distribution of the intensity of the electric field along electric and magnetic diameters of a circular aperture. $ka = 10$.

——————— Exact theory of Andrejewski.
— — — — Edge-current theory: result of numerical integration of eqns. (16) and gradient of eqn. (17).
- — — - — Edge-current theory: numerical evaluation of eqns. (24) and (25) respectively.
— · — · — Sommerfeld half-plane theory. *E*- and *H*-polarizations in (*a*) and (*b*) respectively
- - - - - - - Approximate theory of Bekefi.

(*a*) and (*c*) Along magnetic diameter.
(*b*) and (*d*) Along electric diameter.

On the electric ($x = 0$) and magnetic ($y = 0$) diameters, these expressions reduce to

$$\left.\begin{aligned} E_x(0, y, 0) &= 0 \\ E_y(0, y, 0) &= 1 - \frac{\varepsilon^{-jka\alpha}}{\sqrt{2\alpha}}[J_0(ky/\alpha) + J_2(ky/\alpha)] + 0\left(\frac{1}{ka}\right) \end{aligned}\right\} . \quad (24)$$

and

$$\left.\begin{aligned} E_x(x, 0, 0) &= 0 \\ E_y(x, 0, 0) &= 1 - \frac{\varepsilon^{-jka}}{\sqrt{2}}\Big\{J_0(kx) - J_2(kx) \\ &\qquad + \frac{j}{2}\frac{x}{a}[3J_1(kx) - J_3(kx)]\Big\} + 0\left(\frac{1}{ka}\right) \end{aligned}\right\} . \quad (25)$$

(5) DISCUSSION AND CONCLUSIONS

Figs. 2(*a*)–2(*d*) show the axial electric-intensity distribution computed from eqn. (10), for several values of a/λ, in comparison with the experimental results of Andrews,[1] Severin,[2] and Silver, Ehrlich and Held.[3] Fairly close correspondence exists between experiment and theory, even down to $a/\lambda = \frac{1}{2}$. The divergence between the two is appreciable in the immediate neighbourhood of the aperture, at the centre of which the theoretically determined E_y takes the value $1 - 1/\sqrt{2} \exp(-jka)$.

Agreement for such small apertures as $a/\lambda = \frac{1}{2}$ was not expected from this theory. Andrejewski,[18] in a more recent publication than Reference 6, notes that on the axis certain terms in the exact solution vanish. These are, in fact, the terms determined by the permissible field singularities at the edge of the aperture, and off the axis they determine the field structure especially as regards the influence of the incident polarization. Thus it is to be expected that closest agreement between exact and approximate theories—the latter not necessarily satisfying an "edge condition"—will be obtained on the axis. In this same publication Andrejewski demonstrates the variation with ka of the electric-field intensity at the centre of the aperture in the range $0 \leqslant ka \leqslant 10$. This is found to agree well with the corresponding values from the edge-current theory for $ka \gtrsim 4$, but the two diverge widely for smaller values of ka. The apparent validity of the edge theory for small apertures is probably largely due to the above characteristic possessed by the axis.

The complete lack of agreement in the case $a/\lambda = 5$ was also unexpected, and may be due, on the experimental side, to non-uniformity of the incident wavefront across this large aperture. Recently Ehrlich *et al.*,[19] recognizing that their incident wave is spherical, suggested this as a possible cause for the behaviour of the axial field for $a/\lambda = 5$. If R_0 and R_E are the distances from source to centre and edge respectively, of the aperture, $R_E - R_0 = a^2/(R_E + R_0)$. In the case $R_0 \gg a$ (as would be expected in the experimental arrangement), $R_E - R_0 \simeq a^2/2R_0$. Thus the phase difference between the incident wave at the centre and edge of the aperture varies as a^2, and an effect negligible for $a/\lambda = 3$ [Fig. 2(*c*)] might be quite noticeable for $a/\lambda = 5$ [Fig. 2(*d*)].

The present theory may be simply modified for an incident spherical wave if R_0/a is assumed large enough, so that E^i is still essentially parallel to the y-axis over the entire aperture. In general, the modified theory predicts displaced maxima and minima, and smaller amplitude for $|E_y|^2$ in relation to predictions of the incident plane-wave theory. Little quantitative information can be obtained without knowledge of R_0.

Figs. 3(*a*)–3(*j*) illustrate the variation of the electric intensity along electric and magnetic diameters of circular apertures with radii $\frac{1}{2}$, 1, $\frac{3}{2}$, 2 and 4 wavelengths, computed from eqns. (24) and (25), and compared with the measurements of Andrews,[1] Silver *et al.*,[3] and Bekefi and Woonton.[4] Since eqns. (24) and (25) are valid only for small values of r_p/a, the numerical calculations are restricted to the range $r_p/a \lesssim 0{\cdot}7$. Again the difference between theoretical and experimental results is appreciable at $x = y = 0$.

The results of a numerical integration of eqns. (16) and the gradient of eqn. (17), for $ka = 10$ is compared in Figs. 4(*a*) and 4(*b*) with the exact theoretical values of Andrejewski.[6] A comparison is made, in Figs. 4(*c*) and 4(*d*) between Andrejewski's results, the approximate theory of Bekefi,[11] and numerical evaluation of eqns. (24) and (25), all for $ka = 10$. Close correspondence between the exact and edge theories is obtained except near the edge of the aperture. In this region, as seen in Figs. 4(*a*) and 4(*b*), the field is similar to that of a half-plane with edge parallel to the adjacent edge of the rim. For $ka = 10$, the intensity along $x = 0$ given by the edge theory if $y/a \lesssim 0{\cdot}75$ and given by the half-plane field if $y/a \gtrsim 0{\cdot}75$ is within a few per cent of the exact values, while along $y = 0$ the error is somewhat larger.

The effect of the previously mentioned amplitude difference between the edge theory and Andrews's equations is most noticeable at the centre of the aperture. Here, Andrews's theory gives $E_y = 1 - 1/2 \exp(-jka)$, which is also given to the same order in $1/(ka)$ by Bekefi. On the other hand, the edge-current theory, and, to the same order, Frahn's theory based on half-plane fields, give $E_y = 1 - 1/\sqrt{2} \exp(-jka)$. The first terms in Bekefi's solution for the aperture electric field are

$$\left.\begin{aligned} E_x &= -\frac{\sin 2\theta_p}{2}\varepsilon^{-jka}J_2(kr_p) + 0\left(\frac{1}{ka}\right) \\ E_y &= 1 - \frac{\varepsilon^{-jka}}{2}[J_0(kr_p) - J_2(kr_p)\cos 2\theta_p] + 0\left(\frac{1}{ka}\right) \end{aligned}\right\} \quad (26)$$

whilst, for $r_p/a \ll 1$, the edge-current theory gives

$$\left.\begin{aligned} E_x &= -\frac{\sin 2\theta_p}{\sqrt{2}}\varepsilon^{-jka}J_2(kr_p) + 0\left(\frac{1}{ka}\right) \\ E_y &= 1 - \frac{\varepsilon^{-jka}}{\sqrt{2}}[J_0(kr_p) - J_2(kr_p)\cos 2\theta_p] + 0\left(\frac{1}{ka}\right) \end{aligned}\right\} \quad (27)$$

The similarity between eqns. (26) and (27) is immediately obvious. For apertures with diameters equal to an integral number of wavelengths, all available measurements of the intensity at the centre tend to support eqn. (26), but for $ka = 10$, eqn. (27) gives a value much closer to the exact one of Andrejewski. It is also of interest to note that neither the geometrical-optics theory of Bekefi, nor Andrews's equations yields the correct amplitude of the field scattered by a half-plane; both are in error by a factor of $1/\sqrt{2}$.

In the foregoing, the edge theory has been developed to give the axial and aperture fields only. However, it should prove possible to treat some more general cases. For instance, for a point (r_p, θ_p, z_p) situated slightly off the axis, so that ψ is still essentially constant around the aperture, a term dependent on $kr_p \cos(\theta - \theta_p)$ enters into the exponent of the integrand, and Bessel functions result on integration. On the other hand, when r_p is not sufficiently small to ignore this variation of ψ, expansion of the ψ part of the integrand in powers of r_p, and term-by-term integration is possible.

Another possibility is generalization to oblique incidence, but a number of added complications arise. The edge currents may again be deduced by appealing to the theories of Clemmow[20] and Copson[21] on the quasi three-dimensional half-plane problem, and integrals analogous to those in the present case may be developed. The only feasible method for evaluating these seems be by expansion of the integrands in powers of the sine of the gle between the propagation vector of the incident plane wave, l the normal to the aperture.

(6) ACKNOWLEDGMENTS

The author would like to acknowledge with thanks the many pful suggestions and criticisms from Dr. P. C. Clemmow, and receipt of a Special Scholarship from the National Research uncil of Canada, throughout the period of this work.

(7) REFERENCES

) ANDREWS, C. L.: "Diffraction Pattern in a Circular Aperture measured in the Microwave Region," *Journal of Applied Physics*, 1950, **21**, p. 761.

) SEVERIN, H.: "Beugung elektromagnetischer Zentimeterwellen an metallischen Blenden," *Zeitschrift für Naturforschung*, 1946, **1**, p. 487.

) SILVER, S., EHRLICH, M. J., and HELD, G.: "Diffraction of a Plane Electromagnetic Wave by a Circular Aperture and Complementary Obstacle: Part II, Discussion of Experimental Results," *University of California Antenna Laboratory Report*, Series 7, Issue No. 185, September, 1952.

) BEKEFI, G., and WOONTON, G. A.: "Microwave Diffraction Measurements on Circular Apertures. I. An Investigation of the Electric Field," *Eaton Electronics Research Laboratory, McGill University, Technical Report No.* 24, November, 1952.

) MEIXNER, J., and ANDREJEWSKI, W.: "Strenge Theorie der Beugung ebener elektromagnetischer Wellen an der vollkommen leitenden Kreisscheibe und an der kreisförmigen Öffnung im vollkommen leitenden ebenen Schirm," *Annalen der Physik*, 1950, **7**, p. 157.

) ANDREJEWSKI, W.: "Strenge Theorie der Beugung ebenen elektromagnetischer Wellen an der vollkommen leitenden Kreisscheibe und an der kreisförmigen Öffnung im vollkommen leitenden ebenen Schirm. Numerische Ergebnisse," *Die Naturwissenschaften*, 1951, **38**, p. 406.

) BRAUNBEK, W.: "Neue Näherungsmethode für die Beugung am ebenen Schirm," *Zeitschrift für Physik*, 1950, **127**, p. 381.

) BRAUNBEK, W.: "Zur Beugung an der Kreisscheibe," *ibid.*, 1950, **127**, p. 405.

) BRAUNBEK, W.: "Zur Beugung an der kreisförmigen Öffnung," *ibid.*, 1954, **138**, p. 80.

) FRAHN, W.: Diplomarbeit, Institut für theoretische Physik, Rheinisch-Westfälische Technische Hochschule, Aachen, 1951.

) BEKEFI, G.: "Diffraction of Electromagnetic Waves by an Aperture in a Large Screen," *Journal of Applied Physics*, 1953, **24**, p. 1123.

) MOULLIN, E. B., and PHILLIPS, F. M.: "On the Current Induced in a Conducting Ribbon by the Incidence of a Plane Electromagnetic Wave," *Proceedings I.E.E.*, Monograph No. 26 R, March, 1952 (**99**, Part IV, p. 137).

) BAKER, B. B., and COPSON, E. T.: "The Mathematical Theory of Huygens' Principle" (Oxford University Press, 1950).

(14) STRATTON, J. A.: "Electromagnetic Theory" (McGraw-Hill, 1941).

(15) VAN DE HULST, H. C.: "On the Attenuation of Plane Waves by Obstacles of Arbitrary Size and Form," *Physica*, 1949, **15**, p. 740.

(16) LEVINE, H., and SCHWINGER, J.: "On the Theory of Electromagnetic Wave Diffraction by an Aperture in an Infinite Plane Conducting Screen," *Communications on Pure and Applied Mathematics*, 1950, **3**, p. 355.

(17) STENZEL, H.: "Über die Berechnung des Schallfeldes unmittelbar vor einer kreisförmigen Kolbenmembran," *Annalen der Physik*, 1942, **41**, p. 245.

(18) ANDREJEWSKI, W.: "Die Beugung elektromagnetischer Wellen an der leitenden Kreisscheibe und an der kreisförmigen Öffnung im leitenden ebenen Schirm," *Zeitschrift für angewandte Physik*, 1953, **5**, p. 178.

(19) EHRLICH, M. J., SILVER, S., and HELD, G.: "Studies of the Diffraction of Electromagnetic Waves by Circular Apertures and Complementary Obstacles: The Near-Zone Field," *Journal of Applied Physics*, 1955, **26**, **p. 336**.

(20) CLEMMOW, P. C.: "A Method for the Exact Solution of a Class of Two-Dimensional Diffraction Problems," *Proceedings of the Royal Society*, A, 1951, **205**, p. 286.

(21) COPSON, E. T.: "Diffraction by a Plane Screen," *ibid.*, 1950, **202**, p. 277.

(8) APPENDIX

Evaluation of div Π.

From eqns. (16), (18) and (19), and with the variable of integration θ rather than β,

$$\left.\begin{aligned} \Pi_x(x, y, 0) &= \frac{a}{\sqrt{2\pi k^2}}\int_0^{2\pi}\frac{\varepsilon^{-jk\rho}}{\rho}\sin\theta\cos\theta d\theta \\ \Pi_y(x, y, 0) &= -\frac{a}{\sqrt{2\pi k^2}}\int_0^{2\pi}\frac{\varepsilon^{-jk\rho}}{\rho}\cos^2\theta d\theta \end{aligned}\right\} \quad . \quad (28)$$

where $\rho^2 = x^2 + y^2 + a^2 - 2a(x\cos\theta + y\sin\theta) \quad . \quad (29)$

Then div $\Pi(x, y, 0) =$

$$\frac{a}{\sqrt{2\pi k^2}}\int_0^{2\pi}\left(\sin\theta\cos\theta\frac{\partial\rho}{\partial x} - \cos^2\theta\frac{\partial\rho}{\partial y}\right)\frac{d}{d\rho}\left(\frac{\varepsilon^{-jk\rho}}{\rho}\right)d\theta$$

Since $\dfrac{\partial\rho}{\partial x} = \dfrac{x - a\cos\theta}{\rho}, \; \dfrac{\partial\rho}{\partial y} = \dfrac{y - a\sin\theta}{\rho}$

therefore

$$\text{div}\,\Pi(x, y, 0) = \frac{a}{\sqrt{2\pi k^2}}\int_0^{2\pi}\cos\theta\frac{(x\sin\theta - y\cos\theta)}{\rho}\,\frac{d}{d\rho}\left(\frac{\varepsilon^{-jk\rho}}{\rho}\right)d\theta$$

But $\dfrac{d\rho}{d\theta} = \dfrac{a(x\sin\theta - y\cos\theta)}{\rho}$

Hence $\text{div}\,\Pi(x, y, 0) = \dfrac{1}{\sqrt{2\pi k^2}}\displaystyle\int_0^{2\pi}\cos\theta\frac{d}{d\theta}\left(\frac{\varepsilon^{-jk\rho}}{\rho}\right)d\theta \quad . \quad (30)$

An integration by parts now yields eqn. (17).

621.317.326

Monograph No. 153 M
Nov. 1955

PEAK-VOLTAGE MEASUREMENTS OF STANDARD IMPULSE VOLTAGE WAVES

By A. AKED, B.Sc., Graduate.

(*The paper was first received* 19*th May, and in revised form* 2*nd August,* 1955. *It was published as an* Institution Monograph *in November,* 1955.)

SUMMARY

The frequency spectrum of a 1/50 microsec impulse wave is obtained by the use of the Fourier transform. The corresponding frequency spectrum of the output from a simple capacitance potential divider, suitable for use up to 300 kV, is calculated from the input frequency spectrum for a 1/50 microsec wave and the frequency response of the divider, measured up to about 3 Mc/s. From this output frequency spectrum the output at a time of 1 microsec is calculated, and thus the divider ratio, suitable for peak-voltage measurements, is estimated. Consideration is given to errors due to the limited frequency range of the measurements, and the accuracy of the measurement of the divider ratio is estimated to be better than ±1%.

(1) INTRODUCTION

The problem of faithfully recording very fast transients has been the subject of a number of investigations during recent years.[1,2,3,4,5] The main considerations have been to analyse the defects of existing types of impulse voltage dividers[1,2,3] and to attempt to produce distortionless dividers.[4,5] These investigations have required knowledge of the frequency spectra of impulse waves and of the frequency responses of potential dividers. In these investigations interest has centred mainly on attempts to obtain frequency responses which remain level up to the highest frequencies required to give undistorted records. If the frequency-response curve of a potential divider is not level, the output will be distorted. The size of the distortion can be calculated, for any time t, from the frequency spectrum of the applied impulse and the frequency response of the divider. This method has been applied to determine the appropriate ratio of a simple capacitance potential divider for peak-voltage measurements on standard impulse voltage waves.

Measurements of the peak values of standard 1/50 microsec impulse voltages up to 300 kV were required during investigation of the impulse breakdown characteristics of uniform-field spark gaps. The impulse voltages were recorded by a high-speed cathode-ray oscillograph and a simple capacitance potential divider, as shown in Fig. 1. The capacitance divider C and the resistor R produce the required wavefront of 1 microsec nominal duration. The resistors necessary to give a wavetail of 50 microsec were incorporated in the generator, and also acted as charging resistances. Tripping of the generator was controlled, so that a delay cable in the oscillograph circuit was unnecessary.

(2) THEORY

The ideal impulse wave, assumed to be applied to the divider, was represented by a simple mathematical function. This function was suitable for transformation by the Fourier transform into a frequency spectrum showing the relative amplitudes and phase angles of the continuously distributed frequency components. The transfer function of the divider, giving the relationship between the output and input, was measured for a wide range of frequencies. The frequency spectrum of the output is the product of the input frequency spectrum and the transfer function. From this output frequency spectrum, the output at time $t = 1$ microsec was calculated by the use of the inverse Fourier transform, and thus the divider ratio suitable for peak-voltage measurements was obtained.

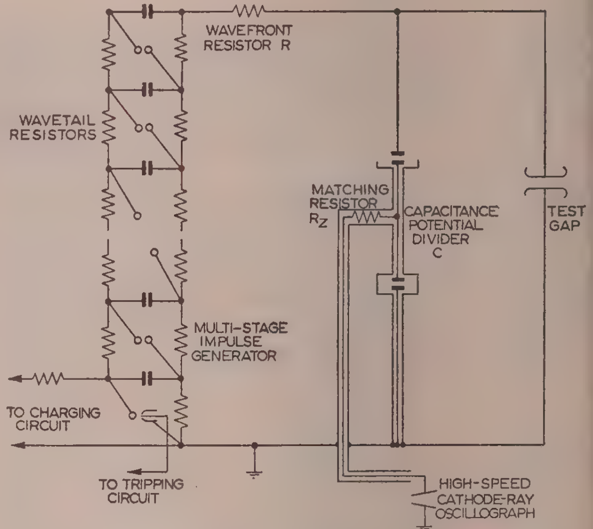


Fig. 1.—Multi-stage impulse generator and recording circuit.

(3) EQUATION OF THE THEORETICAL IMPULSE WAVE

The impulse wave was a nominal 1/50 microsec wave.[6] For analytical purposes it was assumed that the wave could be represented by an equation of the form

$$v_t = V(\varepsilon^{-a_1 t} - \varepsilon^{-a_2 t}) \quad . \quad . \quad . \quad . \quad . \quad (1)$$

the peak of the wave being at $t_1 = 1$ microsec and half the peak value occuring on the tail at $t_2 = 50$ microsec. The constants in the equation are $a_1 = 0{\cdot}0142$, $a_2 = 6{\cdot}073$ and $V = 1{\cdot}0167 v_{t1}$. The equation of the theoretical 1/50 microsec wave is therefore

$$v_t = 1{\cdot}0167 v_{t1}(\varepsilon^{-0{\cdot}0142t} - \varepsilon^{-6{\cdot}073t}) \quad . \quad . \quad . \quad (2)$$

where t is in microseconds.

(4) FOURIER TRANSFORM

The magnitudes of the components of a time function f(t) are given by the Fourier transform F(ω).

$$\text{In general,} \quad F(\omega) = \int_{-\infty}^{\infty} f(t)\varepsilon^{-j\omega t}dt \quad \text{for } \omega \neq 0 \quad . \quad . \quad . \quad (3)$$

$$\text{and} \qquad F(0) = \int_{-\infty}^{\infty} f(t)dt \quad \text{for } \omega = 0 \quad . \quad . \quad . \quad . \quad (4)$$

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Aked is at the Royal Technical College, Glasgow.

'or the function $f(t) = \varepsilon^{-at}$

$$\left.\begin{aligned} F(\omega) &= \int_{-\infty}^{\infty} \varepsilon^{-at}\varepsilon^{-j\omega t}dt \\ &= 1/(a + j\omega) \\ &= R_a\angle\theta_a \end{aligned}\right\} \text{for } \omega \neq 0 \quad . \quad . \quad (5)$$

$$\left.\begin{aligned} F(0) &= \int_{-\infty}^{\infty} \varepsilon^{-at}dt \\ &= 1/a \\ &= R_0 \end{aligned}\right\} \text{for } \omega = 0 \quad . \quad . \quad (6)$$

The R_a and θ_a curves to a base of ω/a are shown in Fig. 2. 'he transform of the impulse wave, represented by $\varepsilon^{-a_1t} - \varepsilon^{-a_2t}$, ; the difference of the transforms of ε^{-a_1t} and of ε^{-a_2t}. For

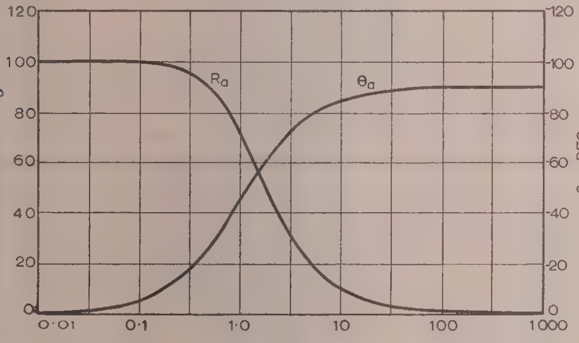


Fig. 2.—Fourier transform of ε^{-at} to a base of ω/a.

$F(\omega) = 1/(a + j\omega) = R_a\angle\theta_a$. $F(0) = 1/a = R_0\angle\theta_0$.

he 1/50 microsec wave the frequencies at which R_{a1} and R_{a2} all to 1% of their maximum values are 226 kc/s and 9·67 Mc/s, espectively. The maximum value of R_{a2} is only 0·23% of the eak value of R_{a1}.

During the preparation of the paper the frequency spectra of standard impulse waves were given by Miles.[7] His analysis roduced an expression similar to that given above, but the values of a_1 and a_2 were slightly different and gave small errors on the eak value and on the value at 50 microsec. In concluding that he significant frequency range is 0–0·1 Mc/s for a full /50 microsec wave, Miles ignored wavefront components which, lthough apparently small in magnitude, are very significant in letermining the shape of the impulse wave. These components lso make a small contribution to the peak value of the output 'oltage. The range, to include wavefront components down to % of their maximum value, would have to extend to nearly 0 Mc/s.

5) FREQUENCY RESPONSE OF THE POTENTIAL DIVIDER

The frequency spectrum of the 1/50 microsec impulse wave hows that a recording system would require a constant response ip to some 10 Mc/s if it were to record faithfully the shape of he impulse wave. Since the main consideration was to measure eak values of 1/50 microsec impulse waves, the frequency esponse of the divider circuit was measured up to about 3 Mc/s only.

6) THE CONSTRUCTION OF THE POTENTIAL DIVIDER

The simple capacitance potential divider had an upper-arm apacitance of approximately 300 μμF, and five lower arms were used to cover the range 30–300 kV. Each lower arm consisted of five to ten capacitors connected in parallel between two brass discs, with the cable terminating resistor placed in the centre of the ring of capacitors, as shown in Fig. 3. This design reduced the self-inductance to a very low value, and the symmetry prevented spurious oscillations occurring between the individual capacitors.

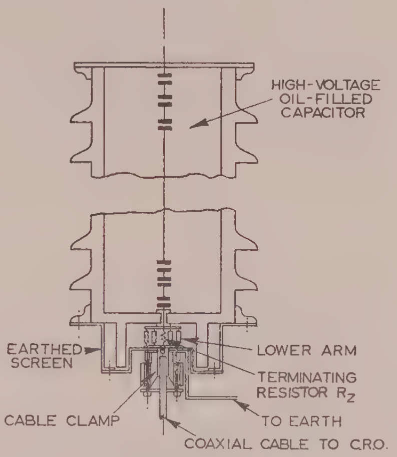


Fig. 3.—Diagram showing construction of the capacitance divider.

(7) MEASUREMENTS AT 50 c/s

The ratio and phase-shift of the potential divider were measured at 50 c/s by the high-voltage bridge circuit shown in Fig. 4. To check the bridge, the ratio and phase-shift of a standard potential divider, using a 100 μμF air standard capacitor as the high-voltage arm and a 0·03 μF shielded capacitor as the lower arm, were measured and the results agreed with those obtained by calculation from previous calibrations of the two capacitors.

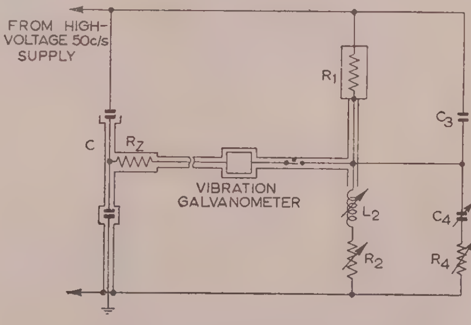


Fig. 4.—High-voltage bridge circuit.

C Capacitance potential divider.
R_1 Standard 1-megohm high-voltage resistor.

(8) MEASUREMENTS AT HIGH FREQUENCIES

Fig. 5 shows the bridge circuit used to measure the transfer function of the divider at frequencies between 50 and 400 kc/s. The standard branch of the bridge contained two radio-frequency decade resistors, R_1 and R_2, and a variable air standard capacitor C_2. Above 400 kc/s R_1 was replaced by a small air capacitor in series with a resistor, and R_2 and C_2 by a variable air capacitor; this circuit was used for measurements up to 2·7 Mc/s. The output from the potential divider was fed directly to one side of the detector, the other side being switched to one of the points 1, 2 or 3.

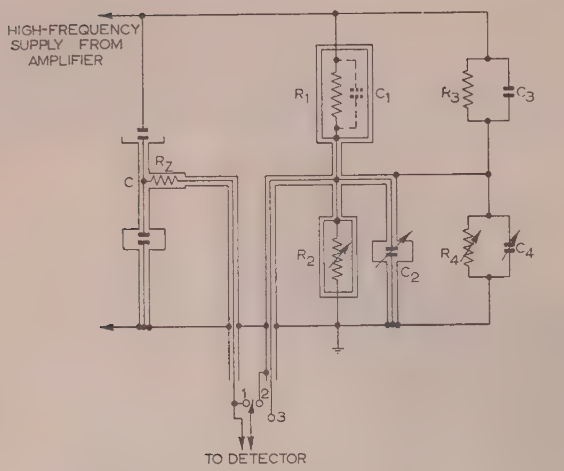


Fig. 5.—High-frequency bridge circuit.

C Capacitance potential divider.
R_1 and R_2 Radio-frequency decade resistors.
C_2 Air standard capacitor.

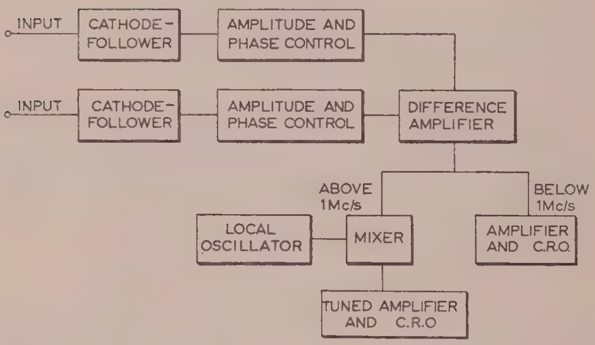


Fig. 6.—Block schematic of the detector in the high-frequency bridge circuit.

A block schematic of the detector is given in Fig. 6. Cathode-follower input circuits were used to give the high input impedance required. The amplitude and phase-shift circuits were used in standardizing the detector. The signal from the capacitance divider was applied to both sides of the detector, and the amplitudes and phases were adjusted until zero output was obtained. The difference amplifier was a tuned-anode pentode, with the signals applied to the cathode and the control grid. A detection sensitivity of 0·1% was achieved.

In the resistance bridge the major source of error is the stray capacitance C_1 in parallel with R_1. If the divider under test has zero phase-shift, then at balance R_1C_1 equals R_2C_2, and the bridge can be balanced for varying values of R_1 by altering R_2 only. This condition was obtained using the standard capacitance divider, and thus C_1 was estimated.

A typical transfer function obtained with the simple capacitance divider is shown in Fig. 7. A simple equivalent circuit with a frequency-response curve approximating closely to the experimental curves could not be found. However, if both the upper and lower arms are represented by a series connection of R, L and C, a rising characteristic with increasing phase-shift is obtained if the natural frequency of the low-voltage arm is higher than that of the high-voltage arm. The natural frequency of the upper arm of the divider was approximately 6 Mc/s.

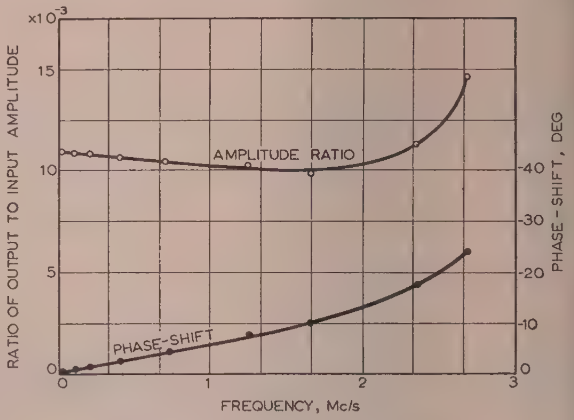


Fig. 7.—Typical divider transfer function.

○ Ratio of output to input amplitude.
● Phase-shift.

(9) INVERSE FOURIER TRANSFORM

At a given frequency, f, the component of the input wave is

$$R_{a1}\,\angle\theta_{a1} - R_{a2}\,\angle\theta_{a2}.$$

If we consider the component $R_{a1}\angle\theta_{a1}$, the divider transfer function at this frequency is $K\angle\phi$, and therefore the corresponding component of the output is $R_{a1}K\angle\phi + \theta_{a1} = G$.

Let $G = A_1 + jB_1$ where A_1 is an even function and B_1 is an odd function, the phase-shift being negative at negative frequencies.

The magnitude of the output at time t is

$$v_t = \frac{1}{2\pi}\int_{-\infty}^{\infty}(A_1 + jB_1)\varepsilon^{j\omega t}d\omega$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}(A_1 + jB_1)(\cos\omega t + j\sin\omega t)d\omega$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}(A_1\cos\omega t - B_1\sin\omega t)d\omega + \frac{j}{2\pi}\int_{-\infty}^{\infty}(A_1\sin\omega t + B_1\cos\omega t)d\omega$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}(A_1\cos\omega t - B_1\sin\omega t)d\omega \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

For $t < 0$, $v_t = 0$

Therefore

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}(A_1\cos\omega t + B_1\sin\omega t)d\omega = 0$$

Therefore $$\int_{-\infty}^{\infty}A_1\cos\omega t\,d\omega = -\int_{-\infty}^{\infty}B_1\sin\omega t\,d\omega$$

Thus $$v_t = \frac{2}{2\pi}\int_{-\infty}^{\infty}A_1\cos\omega t\,d\omega$$

$$= \frac{4}{2\pi}\int_{0}^{\infty}A_1\cos\omega t\,d\omega$$

$$= 4\int_{0}^{\infty}A_1\cos 2\pi ft\,df \quad . \quad . \quad . \quad . \quad (8)$$

nd for the component a_2

$$v_t = 4\int_0^\infty A_2 \cos 2\pi ft\, df \quad . \quad . \quad . \quad . \quad . \quad (9)$$

he difference of these two functions gives the output at time t.

In determining the output of the potential divider at = 1 microsec the above integrations were performed graphically.

(10) ESTIMATE OF ERRORS DUE TO LIMITATION OF FREQUENCY RANGE

The graphical integration was carried out over the range of requencies of the measurements. Theoretically the integration hould be continued to infinity, and limitation of the frequency ange introduces errors. These errors were calculated by ssuming that the impulse wave was passed through an ideal ircuit having a transfer function of $1\angle 0$ from zero to a frequency and thereafter a transfer function of zero. The output from his circuit at 1 microsec was found by graphical integration and compared with the actual input at that time. This gave an stimate of the possible error due to measurements not having been made at frequencies above f.

The estimated error for an input of $\varepsilon^{-0.0142t}$ and a cut-off requency of 1 Mc/s is only 0·16%. Measurements were made up to 3 Mc/s, and thus errors due to the limited frequency range are negligible for this input.

To reduce the error in the estimate of the 1 microsec value of he wavefront function $\varepsilon^{-6.073t}$ to a very small value, the cut-off requency would require to be about 10 Mc/s. For this cut-off requency the true value of the function at 1 microsec is 0·23% of its maximum, and the estimated value is 0·16%. The measurements were, however, only made up to 3 Mc/s, and an estimate must therefore be made of the possible errors from the divider esponse to frequencies above 3 Mc/s.

11) POSSIBLE ERRORS FROM THE DIVIDER RESPONSE O FREQUENCIES ABOVE THOSE OF THE MEASUREMENTS

For an input of fixed amplitude, the output from the divider ises at the higher frequencies owing to the approach to series esonance in the upper arm. The greater error in the calculated output will be due to the components of the wavefront xponential. Measurements could not be made above 3 Mc/s, and attempts to extrapolate the curves by determining an quivalent circuit were unsuccessful. The order of magnitude of the possible error was investigated by assuming a simple quivalent circuit containing series resistance, inductance and apacitance in the upper arm, and a pure capacitance in the ower. This circuit was arranged to have an output, at resonance, of ten times the output calculated from the ratio of the apacitances of the two arms. The resonant frequency was pproximately the same as that occurring after the breakdown of a spark-gap connected to the divider by a very small loop. The error produced by integrating to 3 Mc/s instead of to 10 Mc/s was only 3%. For the ideal impulse wave considered the omponents at the higher frequencies will be larger than the orresponding components of an actual impulse wave, owing o the effect of the circuit inductance on the initial rate of rise of the voltage. Thus the effect of the peaking of the output t the higher frequencies will be reduced. The effect of the peaking would be to produce oscillations on the recorded impulse waveshape, but the smooth waves actually recorded show that the amplitude of these oscillations is negligibly small.

(12) ACCURACY OF THE PEAK-VOLTAGE MEASUREMENTS

An estimation of the accuracy of the peak voltage measurements must include consideration of the accuracy of the measurement of the divider ratio, and also of possible errors occurring in the associated oscillography. The results of the tests on the standard divider and consideration of the smoothness of the recorded impulse waves indicate that the accuracy of the divider-ratio measurements is better than ±1%. The errors involved in calibrating the cathode-ray oscillograph and measuring the oscillograms are estimated at about ±1%. The overall accuracy of the peak voltage measurements for 1/50 microsec waveforms is therefore within ±2%.

(13) CONCLUSIONS

The simple capacitance potential divider, with the self-inductance of the lower arm reduced to a minimum, has been entirely satisfactory for peak-voltage measurements on standard 1/50 microsec impulse waves. The frequency response of the divider used was not level, but the method of analysis enabled the ratio required for the peak voltage measurements to be calculated to within ±1%. For the theoretical impulse wave the components above a frequency of about 200 kc/s are all less than 1% of the maximum value of R_{a1}, but they cannot be neglected; and for peak voltage measurements on a 1 microsec wavefront, it would be desirable to know the frequency response of the divider up to about 10 Mc/s. If the waveform recorded is smooth at the peak, a knowledge of the response up to about 3 Mc/s only is quite sufficient for an overall accuracy of ±2%.

(14) ACKNOWLEDGMENTS

The author wishes to express his thanks to the trustees of the Sir James Caird Travelling Scholarship Trust for their support and also to thank Prof. F. M. Bruce of the Royal Technical College, Glasgow, for his encouragement and suggestions.

(15) REFERENCES

(1) ANGELINI, A. M.: "Diviseurs de tension et câbles retardateurs dans l'enregistrement oscillographique, des phénomènes transitoires rapides," *Bulletin Association Suisse des Électriciens*, 1941, **32**, p. 305.

(2) HOHL, H.: "Der Hochspannungteiler beim Kathodenstrahloszillographen," *Archiv für Electrotechnik*, 1941, **35**, p. 663.

(3) BOCHMAN, M., and HYLTEN-CAVALLIUS, N.: "Errors in Measuring Surge Voltages by Oscillograph" (*Technical Achievements of ASEA Research*, 1946, Vesteras, Sweden).

(4) NAZAROV, S. A.: "Distortionless Divider of Impulse Voltages," *Elektrichestvo*, 1948, No. 8, p. 28.

(5) DAWES, C. L., THOMAS, C. H., and DROUGHT, A. B.: "Impulse Measurements by Repeated-Structure Networks," *Transactions of the American I.E.E.*, 1950, **69**, Part 1, p. 571.

(6) British Standard 923: 1940.

(7) MILES, J. G.: "Frequency Spectra of Standard Impulse Waveshapes," *Metropolitan-Vickers Gazette*, 1954, **25**, p. 367.

621-52

Monograph No. 154 M
Nov. 1955

AN INTRODUCTION TO THE ANALYSIS OF NON-LINEAR CONTROL SYSTEMS WITH RANDOM INPUTS

By J. F. BARRETT and J. F. COALES, O.B.E., M.A., Member.

(*The paper was first received* 11*th December*, 1954, *and in revised form* 2nd *August*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in November*, 1955.)

SUMMARY

By way of introduction, the Wiener method of optimization of linear control systems with noisy inputs is briefly set out, and attention is called to the reasons why this method cannot be used directly for systems containing non-linear components. It is then shown that the difficulty with random inputs arises because the probability distribution of the output can be calculated only when that of the error signal is known; but in order to obtain this quantity the distribution of input and output must be combined. Thus it is not, in general, possible to obtain explicit expressions for either the error function or the output. If all the distributions were Gaussian, spectral densities could be used to obtain a solution of the problem, and this is the basis of Burt's approximation (not previously published), which is given in detail. Unfortunately, when the input of a non-linear component is Gaussian, the output will be non-Gaussian, but in some cases it is possible to make the necessary approximation.

The mathematical justification for this approximation and for Booton's approximation is given in the Appendix. Booton's approximation consists in dividing the non-linear characteristic into a linear part and what he calls "the distortion factor." The slope of the linear part is adjusted to give the best fit on a mean-square-error basis, and the distortion factor is then neglected.

A method of obtaining an explicit solution to any required degree of accuracy by approximating to the non-linear characteristic by a number of linear domains is given in the Appendix.

Finally, the possible use of topological methods to determine the stability of non-linear systems with random inputs is discussed, and it is shown that for useful control systems the phase-plane diagram of error for no input can almost certainly be used in the design of non-linear systems, provided that the bounds of input and output and their derivatives can be determined.

INTRODUCTION

The study of non-linear control systems with random inputs is necessary because for them the principle of superposition does not apply. In a linear system, if the performance is known in terms of a step-function, it can be converted into the frequency response, and conversely. Further, the work of Wiener[1] has shown how the performance for a random input can be calculated from the frequency response, provided that the frequency spectrum or autocorrelation function of the input is known. Thus the design of control systems to fulfil predetermined criteria under conditions of random input has become possible for linear systems, even in the presence of unwanted disturbances or noise.

In non-linear systems with random inputs, however, the probability distribution of the input has to be combined with that of the output in order to obtain the probability distribution of the error function on which the output depends. When this is attempted it is found that simultaneous integral equations are obtained which cannot be solved by simple means. It is the object of the paper to show how these integral equations arise and to review the methods of approximation which have been used to date. In the Appendix these approximations have been formally justified by the authors under certain conditions, the nature of which is then evident. Since a great deal of the analysis of control systems is concerned mainly with stability, it is the presence of a random forcing function (the input) which makes a rigid solution difficult or impossible to attain in many cases. The possibility of using topological methods to investigate stability is also discussed, and it is shown that for second-order systems a phase-plane method will yield useful information concerning stability when the input and output and their derivatives are bounded as in all practical systems. It is evident that the arguments employed can be extended for systems of higher order.

(1) LINEAR SYSTEMS WITH RANDOM INPUTS

The basis of this analysis is that if the input to a linear system, having impulse response $h_0(t)$, is a function of time $x(t)$, the output $y(t)$ is the sum of the responses of the system to all the unit impulses which go to make up $x(t)$. This is expressed by the convolution integral

$$y(t) = \int_0^\infty h_0(\tau)x(t-\tau)d\tau \quad . \quad . \quad . \quad . \quad . \quad (1)$$

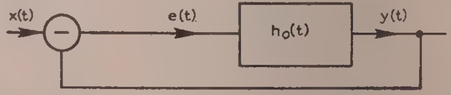


Fig. 1.—Simple linear control system with random input.

In a linear control system (Fig. 1) in which the impulse response of the open loop is $h_0(t)$ it follows that

$$\left.\begin{aligned} y(t) &= \int_0^\infty h_0(\tau)e(t-\tau)d\tau \\ e(t) &= x(t) - y(t) \quad . \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (2)$$

so that $y(t) = \int_0^\infty h_0(\tau)[x(t-\tau) - y(t-\tau)]d\tau$

$$= \int_0^\infty h_0(\tau)x(t-\tau)d\tau - \int_0^\infty h_0(\tau)y(t-\tau)d\tau \quad . \quad (3)$$

It is quite easily shown that if eqn. (3) is written

$$y(t) = \int_0^\infty h(\tau)x(t-\tau)d\tau \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Correspondence on Monographs is invited for consideration with a view to publication.
The authors are in the Engineering Laboratory, University of Cambridge.

(t), the impulse response of the closed loop, is given by

$$[H(p)] = \frac{[H_0(p)]}{1 + [H_0(t)]} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

vhere $[H(p)]$ is the Laplace transform of $h(t)$.

It is usually some function of $e(t)$ that will be used as a measure of performance, and it is convenient to write

$$e(t) = x(t) - y(t) = x(t) - \int_0^\infty h(\tau)x(t-\tau)d\tau \quad . \quad . \quad (6)$$

which is an explicit expression for $e(t)$ in terms of $x(t)$ and t, ince $h(t)$ is a known function of t. Further, knowing the spectral lensity of $x(t)$, it is in general possible by the calculus of variaions to find the function $h(t)$ which minimizes any desired unction of $e(t)$. If, however, the system contains an instantaneous non-linear element, $f(e)$, which transforms the error function $e(t)$ into $f(e)$, as in Fig. 2, before it is operated on by he linear network, $h_0(t)$, then

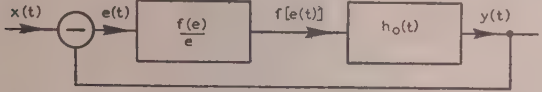


Fig. 2.—Simple non-linear control system with random input.

$$\left.\begin{aligned} y(t) &= \int_0^\infty h_0(\tau)\, f[e(t-\tau)]d\tau \\ e(t) &= x(t) - y(t) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (7)$$

giving
$$x(t) = e(t) + \int_0^\infty h_0(\tau)\, f[e(t-\tau)]d\tau \quad . \quad . \quad . \quad (8)$$

which does not yield an explicit expression for $e(t)$ in terms of $x(t)$ and t.

From the point of view of performance it is interesting to nvestigate some norm of $e(t)$, such as the average value of $|e(t)|$, lenoted by $\overline{|e(t)|}$, or the time mean value of $e^2(t)$ defined by $\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e^2(t)dt$. For the present purpose it will be more convenient to take the time mean value of $e^2(t)$, denoted by $\overline{e^2(t)}$, as the measure of performance.

The square of the error is given from eqns. (2) and (4):

$$[e(t)]^2 = x(t)x(t) - 2x(t)\int_0^\infty h(\tau)x(t-\tau)d\tau$$
$$+ \int_0^\infty\int_0^\infty h(\tau_1)h(\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 \quad . \quad (9)$$

Taking the mean of both sides gives

$$\phi_e(0) = \phi_x(0) - 2\int_0^\infty h(\tau)\phi_x(\tau)d\tau$$
$$+ \int_0^\infty\int_0^\infty h(\tau_1)h(\tau_2)\phi_x(\tau_1 - \tau_2)d\tau_1 d\tau_2 \quad . \quad (10)$$

where $\phi_x(\tau)$ is called the autocorrelation function of $x(t)$ and is defined by

$$\phi_x(\tau) = \overline{x(t)x(t+\tau)} = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x(t)x(t+\tau)d\tau \quad . \quad (11)$$

so that $\phi_x(0) = \overline{x(t)x(t)} = \overline{[x(t)]^2}$ and $\phi_e(0) = \overline{e^2(t)}$. (11*a*)

It is quite easily shown that

$$\int_0^\infty\int_0^\infty h(\tau_1)h(\tau_2)\overline{x(t-\tau_1)x(t-\tau_2)}d\tau_1 d\tau_2$$
$$= \int_0^\infty \phi_h(\tau)[\phi_x(\tau) + \phi_x(-\tau)]d\tau \quad . \quad (12)$$

where $\phi_h(\tau)$ is called the filter correlation function[2] of $h(t)$ defined by

$$\phi_h(\tau) = \int_0^\infty h(t)h(t+\tau)dt \quad . \quad . \quad . \quad . \quad (13)$$

Provided that $x(t)$ is a stationary function of time

$$\phi_x(-\tau) = \phi_x(\tau) \quad . \quad . \quad . \quad . \quad . \quad (14)$$

Eqn. (10) becomes

$$e^2(t) = \phi_e(0) = \phi_x(0) - 2\int_0^\infty h(\tau)\phi_x(\tau)d\tau + 2\int_0^\infty \phi_h(\tau)\phi_x(\tau)d\tau \; . \; (15)$$

which can, in general, be quite easily evaluated when $h(t)$, $\phi_h(\tau)$ and $\phi_x(\tau)$ are known.

It is now evident that, since $\phi_x(\tau)$ is a function of $h(\tau)$ and therefore of $h_0(\tau)$, it is, in general, possible to find a theoretical impulse response $h_0(t)$ which will reduce $\overline{e^2(t)}$ to a minimum; but this may not be physically realizable.

Suppose that noise is present in the system; in a linear system it can, by suitable transformation, always be referred to the input, no matter where it occurs in the system. It is convenient, therefore, to represent these unwanted disturbances (noise) by a function $n(t)$, at the input, which will be taken to be stationary. The wanted signal, also a stationary function of time, will be denoted by $s(t)$. The input, $x(t)$, is now given by

$$x(t) = s(t) + n(t) \quad . \quad . \quad . \quad . \quad . \quad (16)$$

In the simple case shown in Fig. 1, in order that the output may follow the wanted signal $s(t)$ as nearly as possible, the error can be defined by

$$e(t) = s(t) - y(t) \quad . \quad . \quad . \quad . \quad . \quad (17)$$

But
$$y(t) = \int_0^\infty h(\tau)x(t-\tau)d\tau \quad . \quad . \quad . \quad . \quad (4)$$

so that
$$e(t) = s(t) - \int_0^\infty h(\tau)x(t-\tau)d\tau$$

Hence

$$[e(t)]^2 = [s(t)]^2 - 2s(t)\int_0^\infty h(\tau)x(t-\tau)d\tau$$
$$+ \int_0^\infty\int_0^\infty h(\tau_1)h(\tau_2)\overline{x(t-\tau_1)x(t-\tau_2)}d\tau_1 d\tau_2 \quad . \quad (18)$$

and taking the mean of both sides

$$\phi_e(0) = \phi_s(0) - 2\int_0^\infty \phi_{xs}(\tau)h(\tau)d\tau + 2\int_0^\infty \phi_h(\tau)\phi_x(\tau)d\tau \quad . \quad (19)$$

If the closed-circuit impulse response is to give the minimum mean square error, replacement of $h(t)$ by $h(t) + f(t)$ must result in a larger value of $\phi_e(0)$, denoted by N, so that

$$\begin{aligned} N = \phi_s(0) &- 2\int_0^\infty \phi_{xs}(\tau)h(\tau)d\tau \\ &+ \overline{\int_0^\infty\int_0^\infty h(\tau_1)h(\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2} \\ &- 2\int_0^\infty \phi_{xs}(t)f(\tau)d\tau + \overline{\int_0^\infty\int_0^\infty f(\tau_1)h(\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2} \\ &+ \overline{\int_0^\infty\int_0^\infty h(\tau_1)f(\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2} \\ &+ \overline{\int_0^\infty\int_0^\infty f(\tau_1)f(\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2} \quad . \quad . \quad . \quad (20) \end{aligned}$$

$$\begin{aligned} \geqslant \phi_e(0) &- 2\int_0^\infty f(\tau)\phi_{xs}(\tau)d\tau \\ &+ \overline{\int_0^\infty\int_0^\infty f(\tau_1)h(\tau_2)x(t-\tau_1)x(t+\tau_2)d\tau_1 d\tau_2} \\ &+ \overline{\int_0^\infty\int_0^\infty f(\tau_2)h(\tau_1)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2} \quad . \quad . \quad . \quad (21) \end{aligned}$$

since the last integral of eqn. (20) is the ensemble average of $\left[\int_0^\infty f(\tau)x(t-\tau)d\tau\right]^2$, which must be $\geqslant 0$. Since the term neglected is of the second order, the minimizing condition is

$$\begin{aligned} 0 = &- 2\int_0^\infty f(\tau)\phi_{xs}(\tau)d\tau \\ &+ 2\int_0^\infty\int_0^\infty f(\tau_1)h(\tau_2)\overline{x(t-\tau_1)x(t-\tau_2)}d\tau_1 d\tau_2 \quad . \quad (22) \\ = 2&\int_0^\infty f(\tau_1)\left[-\phi_{xs}(\tau_1) + \int_0^\infty h(\tau_2)\overline{x(t-\tau_1)x(t-\tau_2)}d\tau_2\right]d\tau_1 \end{aligned}$$

since the last two terms of eqn. (21) must be equal, because the variables can be interchanged. The minimizing condition for all impulse functions $f(t)$ thus becomes

$$\begin{aligned} \phi_{xs}(\tau_1) &= \int_0^\infty h(\tau_2)\overline{x(t-\tau_1)x(t-\tau_2)}d\tau_2 \quad . \quad . \quad (23) \\ &= \int_0^\infty h(\tau)\phi_{xx}(\tau_1-\tau)d\tau \ (\tau_1 \geqslant 0) \quad . \quad . \quad (24) \end{aligned}$$

(2) CLOSED SYSTEMS INCLUDING AN INSTANTANEOUS NON-LINEARITY

If the methods of Section 1 be applied to the simple non-linear case shown in Fig. 2, it follows from eqn. (7) that

$$\begin{aligned} [e(t)]^2 = [x(t)]^2 - 2x(t)&\int_0^\infty f[e(t-\tau)]h_0(\tau)d\tau \\ &+ \int_0^\infty\int_0^\infty h_0(\tau_1)h_0(\tau_2)f[e(t-\tau_1)]f[e(t-\tau_2)]d\tau_1 d\tau_2 \quad . \quad (25) \end{aligned}$$

and it is immediately seen that this does not yield an explicit expression for $[e(t)]^2$.

Physically the problem is that, if the probability distribution $P(x)$ of the input function $x(t)$ to the system shown in Fig. 2 be known, then, in order to obtain the probability distribution $P(e)$ of the error function $e(t)$, it is necessary to combine $P(x)$ with the probability distribution $P(y)$ of the output $y(t)$. Even if this probability function were known, it would be difficult, and might be impossible, to carry out, but in any case $P(y)$ is dependent on the probability distribution $P(e)$ of $e(t)$. Now even though the non-linear device is instantaneous, the output will be delayed by the linear filter $h_0(t)$ and so it will be necessary to combine two probability distributions of quantities displaced in time.

If all the probability distributions could be taken as Gaussian, it would be possible to work in spectral densities and correlation functions and so obtain an explicit solution, but, unfortunately, even if the error function had Gaussian distribution, the output from the non-linear device, and hence the output $y(t)$, would not have Gaussian distribution, and so the combination with the input would not be possible. Even if it were, it follows that unless the input had some special non-Gaussian distribution the error-function distribution would not be Gaussian and the original assumption would be false.

This does, however, suggest that it might be possible to find a probability distribution of the input which would make the error-function distribution Gaussian; it should then be possible to determine the performance of the system on the basis of a random input having this arbitrary distribution. Since the probability distribution of the input to a system is not usually known with any high degree of accuracy, in many cases this might provide a useful criterion for comparing systems with the same characteristic function but different parameters. Unfortunately, if the characteristic function of the non-linear device is changed, a new input probability distribution will have to be used, but even so, comparison for inputs of constant variance might still be useful.

If the non-linear device is followed by a linear filter with a narrow passband, as the bandwidth is narrowed so the output more nearly approaches Gaussion distribution, no matter what the distribution of the input. In effect, this is the assumption made in describing function methods, and Burt has made use of this in developing a method based on autocorrelation functions and spectral densities which is given below.

(3) BURT'S METHOD* OF OPTIMIZING CERTAIN NON-LINEAR SYSTEMS WITH NOISE

If the autocorrelation function $\phi_x(\tau)$ of the input to an instantaneous non-linear device be known, it is possible to find the autocorrelation function of the output $\phi_\zeta(\tau)$ by integrating the bivariate probability density of $x(t)$ and $x(t+\tau)$ over the domains allowed by the non-linear device. If $x(t)$ and $x(t+\tau)$

* Contained in a privately circulated note dated November, 1951.

e random and correlated variables, each having Gaussian stribution, the joint probability distribution is, by definition,

$$[x(t), x(t+\tau)] = P(x_1, x_2)$$

$$= \frac{(\mu_{11}\mu_{22} - \mu_{12}^2)^{-1/2}}{2\pi} \exp \frac{-\mu_{22}x_1^2 - \mu_{11}x_2^2 + 2\mu_{12}x_1x_2}{2(\mu_{11}\mu_{22} - \mu_{12}^2)} \quad . \quad . \quad . \quad . \quad (26)$$

here

$$\mu_{11} = \overline{x_1^2} = \phi_x(0)$$
$$\mu_{22} = \overline{x_2^2} = \phi_x(0)$$
$$\mu_{12} = \overline{x_1x_2} = \phi_x(\tau)$$

The autocorrelation function of the output of the non-linear evice $\phi_\zeta(\tau)$ is the average value of $\zeta(t)\zeta(t+\tau)$, any one alue of which will be decided by the non-linear characteristic $x)$. If the input values are x_1 and x_2, the output values will be $x_1)$ and $f(x_2)$; the probability of the value $f(x_1)f(x_2)$ will thereore be $P(x_1, x_2)$ and the mean value of $\overline{f(x_1)f(x_2)} = \phi_\zeta(\tau)$ will e given by

$$\phi_\zeta(\tau) = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 f(x_1) f(x_2) P(x_1, x_2)$$

$$= \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 f(x_1) f(x_2) \frac{[\phi_x(0)^2 - \phi_x(\tau)^2]^{-1/2}}{2\pi}$$

$$\exp \frac{-\phi_x(0)(x_1^2 + x_2^2) + 2\phi_x(\tau)x_1x_2}{[2\phi_x(0)^2 - \phi_x(\tau)^2]} \quad . \quad (27)$$

This integral can be evaluated as an infinite series in $\phi_x(\tau)$ as hown in the Appendix, and in some cases the first term alone ives an adequate approximation. Alternatively, in a number f simple cases, some of which have been worked out by Rice,[3] Middleton,[5] etc., the non-linearity can be approximated to by a umber of linear domains as shown in the Appendix, where an n/off device is used as an example.

In the Appendix it is shown that, for an on/off device with Gaussian input, the autocorrelation function, $\phi_\zeta(\tau)$, of the utput is

$$\phi_\zeta(\tau) = x_0^2\left(\frac{\arcsin \rho}{\pi/2}\right) \quad . \quad . \quad . \quad . \quad (28)$$

vhere ρ is the normalized autocorrelation function of the input $\phi_x(\tau)/\phi_x(0)$ and $\pm x_0$ is the output of the on/off device.

It is also indicated how the integral can be evaluated for ther non-linear functions, when they can be approximated by number of linear domains, and it is worth noting that, if the on-linear function is approximated by n linear domains, n^2 ntegrals must, in general, be evaluated.

Thus, either of these methods gives a definite expression for $\phi_\zeta(\tau)$ in terms of $\phi_x(\tau)$.

In general, the probability distribution of the output of a on-linear device will be different from that of the input, and o the spectral density can be obtained from the autocorrelaion function only by assuming it to be of the same shape as hat of the input changed in magnitude alone. This is the ssential approximation made by Burt. However, in many ases this may be a reasonable approximation, and Burt has hown that it is for a linear rectifier; some of Middleton's[5] vork leads to the conclusion that in some cases it certainly is or a limiter. Further, a narrow passband linear filter after the on-linear device, as is assumed when the describing function nethods are used, will tend to reduce the effect of the approxination on the overall system of Fig. 2. The justification for this approximation and the conditions for its validity are given in the Appendix.

If now the complete system shown diagrammatically in Fig. 3

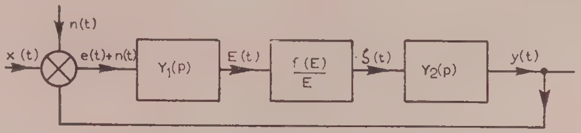


Fig. 3.—Non-linear control system with random input plus noise.

be considered, first of all in the absence of the unwanted disturbance $n(t)$,

$$e = x - y$$
$$L(E) = Y_1(p)L(e) \quad . \quad . \quad . \quad . \quad . \quad (29)$$

where $L(E)$ and $L(e)$ are respectively the Laplace transforms of $E(t)$ and $e(t)$.

It is assumed that the non-linearity does not change the form of the spectral density of E but only reduces its magnitude in the ratio of the output to input mean square values given by

$$\sigma_\zeta^2/\sigma_E^2 = \phi_\zeta(0)/\phi_E(0)$$

and it therefore follows that it can be taken as an amplifier of gain

$$\sigma_\zeta/\sigma_E = \left(\frac{\phi_\zeta(0)}{\phi_E(0)}\right)^{1/2} = R \text{ say}$$

It is now possible to write

$$L(y) = \frac{\sigma_\zeta}{\sigma_E} Y_1(p) Y_2(p) L(e) \text{ (omitting noise)}$$

or

$$L(e) = \frac{L(x)}{1 + RY_1(p)Y_2(p)} \quad . \quad . \quad . \quad . \quad . \quad (30)$$

In order to evaluate R it is necessary to know σ_E, which can be obtained by recourse to the spectral density functions $G(\omega)$, etc. From eqn. (30)

$$G_e(\omega) = \left|\frac{1}{1 + RY_1(j\omega)Y_2(j\omega)}\right|^2 G_x(\omega) \quad . \quad . \quad (31)$$

and so

$$G_E(\omega) = \left|\frac{Y_1(j\omega)}{1 + RY_1(j\omega)Y_2(j\omega)}\right|^2 G_x(\omega) \quad . \quad . \quad (32)$$

$$E^2 = \int_0^\infty G_E(\omega)d\omega = \int_0^\infty \left|\frac{Y_1(j\omega)}{1 + RY_1(j\omega)Y_2(j\omega)}\right|^2 G_x(\omega)d\omega \quad . \quad (33)$$

$R = \left[\frac{\phi_\zeta(0)}{\phi_E(0)}\right]^{1/2}$ can be found in terms of $\sigma_E = [\phi_E(0)]^{1/2}$ by one of the methods given in the Appendix, and eqn. (33) can then be solved for σ_E when G_x is known. $G_E(\omega)$ is immediately obtainable from σ_E and hence $G_e(\omega) = G_E(\omega)/Y_1(j\omega)^2$, as σ_E has been determined, and the value of R is known; hence $G_y(\omega)$ can also be determined. By the relationship $\sigma^2 = \int_0^\infty G d\omega$ the variance of both $e(t)$ and $y(t)$ can be calculated.

If now the input to the system be considered as a slowly changing function of time, $x(t)$, such that the system can quite easily follow this input except for an acceleration lag, and that rapid unwanted fluctuations denoted by $n(t)$ are superimposed on the input function $x(t)$, then if the mean value of $x(t)$ be taken as $\bar{x}$ it follows that

$$\left.\begin{aligned} x(t) &= \bar{x} + n(t) \\ \text{and} \qquad e(t) &= \bar{x} + n(t) - \bar{y} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (34)$$

where $\bar{x} - \bar{y}$ is sensibly constant and dependent only on the lags of the system, $\bar{y}$ being the mean steady state value of the output.

By writing $\quad e = \bar{e} + e(t)$ where $\bar{e} = \bar{x} - \bar{y}$ . . . (35)

it follows that $\quad L(E) = L[\bar{E} + E(t)] = Y_1(p)L[\bar{e} + e(t)]$. (36)

Now for the system as a whole

$$L(E) = Y(p)L(e) = \frac{Y_1(p)}{1 + RY_1(p)Y_2(p)}L(x)$$

$$= \frac{Y_1(p)}{1 + RY_1(p)Y_2(p)}[\bar{x} + L(n)] \quad . \quad . \quad (37)$$

from which, as before, $\bar{E}$ and $E(t)$ can be calculated, provided that $\phi_n(\tau)$, the autocorrelation function of the noise, be known:

$$\bar{E} = \frac{Y_1(0)}{1 + RY_1(0)_1 Y_2(0)}\bar{x} \text{ and } L(E) = \frac{Y_1(p)}{1 + RY_1(p)Y_2(p)}L(n) \text{ . (38)}$$

If the amplitude distribution of the unwanted fluctuations $E(t)$ at the input to the non-linearity is $P[E(t)]$, the amplitude distribution of E will be $P[E(t) - \bar{E}]$ and this must be used in obtaining the two-way probability distribution $P[E_1, E_2]$ of eqn. (26). It will be seen that this results in a different value for $\phi_\zeta(\tau)$ of eqn. (27) and hence for R in terms of $\bar{E}$ and $[\phi_E(0)]^{1/2}$. In this way a modified transfer function for the whole system is obtained in terms of the mean value of E and the variance of E, $\sigma_E = [\phi_E(0)]^{1/2}$. From eqns. (37) and (38) the values of $\bar{E}$ and $[\phi_E(0)]$ can be obtained in terms of p ($= j\omega$), $\bar{x}$ and $G_y(\omega)$, and hence the new transfer function can be determined in the form of a frequency-response function.

(4) BOOTON'S METHOD

Booton,[6] on the other hand, has approached the problem from the point of view of distortion by the non-linear device and by a quasi-linearization approximation. Thus, if the non-linear device is defined by

$$y = f(x)$$

where y is the instantaneous output for an instantaneous input x, he writes

$$y(t) = Kx(t) + x_H(t) \quad . \quad . \quad . \quad . \quad (39)$$

where K is the equivalent gain and $x_H(t)$ is called the "distortion factor." The equivalent gain K must depend on some characteristic of the input function and should be chosen so that the approximation is an optimum in some respect, such as that the mean square error is a minimum. The linearizing approximation then consists in choosing the best value of K when $x_H(t)$ is ignored.

(4.1) The Equivalent Gain

For a particular value of the input, x, the squared error is

$$(y - Kx)^2 = [f(x) - Kx]^2 \quad . \quad . \quad . \quad (40)$$

and the mean square error M is obtained by averaging the right-hand side

$$M = \int_{-\infty}^{\infty}[f(x) - Kx]^2 P(x)dx \quad . \quad . \quad . \quad . \quad (41)$$

where $P(x)$ is the first probability density function of $x(t)$. This expression can be expanded to give

$$M = \int_{-\infty}^{\infty} f(x)f(x)P(x)dx - 2K\int_{-\infty}^{\infty} xf(x)P(x)dx$$

$$+ K^2\int_{-\infty}^{\infty} x^2P(x)dx \quad . \quad (42)$$

Differentiation with respect to K gives

$$K = \int_{-\infty}^{\infty} xf(x)P(x)dx \bigg/ \int_{-\infty}^{\infty} x^2P(x)dx \quad . \quad . \quad . \quad (43)$$

In this expression the denominator is the mean square value of the input.

In the Appendix, $K^2\phi_x(\tau)$ is obtained as the first term of the expansion of $\phi_\zeta(\tau)$ which indicates the mathematical nature of Booton's approximation.

Booton gives as an example the case of a limiter with an input having Gaussian distribution about zero, but he does not take noise into account.

(4.2) Application of Booton's Method

Booton and others have applied this linear approximation to a servo mechanism with a limiter,[10] under Gaussian input. The method will apply equally to any servo mechanism with a single amplitude-sensitive non-linear element of the type shown in Fig. 4A. The system analysed is the corresponding linear system in Fig. 4B, under the same Gaussian input distribution.

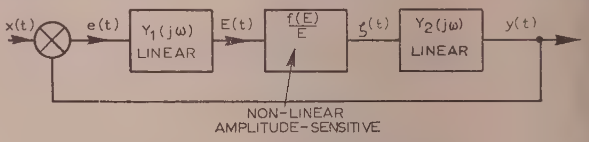


Fig. 4A.—Non-linear control system with random input.

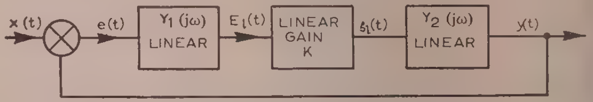


Fig. 4B.—Equivalent linear control system to Fig. 4A.

Since this is linear, the signal in all parts of the system is Gaussian.[11] The gain K is varied until the standard deviation of its input corresponds to the value of K calculated in the previous Section. There may be several values. The investigation of the variation of this quantity σ with input powers can be carried out and is assumed to approximate to the values for the non-linear case.

If $Y_1(j\omega)$ and $Y_2(j\omega)$ are the frequency response functions of the linear parts of the circuit,

$$\overline{y_l} = KY_2(j\omega)\overline{E_l} = KY_1(j\omega)Y_2(j\omega)\overline{e_l} \quad . \quad . \quad (44)$$

where the bar denotes a Fourier transform for a finite signal. But

$$\bar{x} = \bar{y}_l + \bar{e}_l$$

Therefore
$$\overline{e_l} = \frac{\bar{x}}{1 + KY_1(j\omega)Y_2(j\omega)} \quad . \quad . \quad . \quad . \quad (45)$$

and
$$G_{el}(\omega) = \frac{G_x(\omega)}{|1 + KY_1(j\omega)Y_2(j\omega)|^2} \quad . \quad . \quad . \quad (46)$$

Therefore
$$\sigma_e^2 = \int_{-\infty}^{\infty} G_e(\omega)d\omega = \int_{-\infty}^{\infty} \frac{G_x(\omega)}{|1 + KY_1(j\omega)Y_2(j\omega)|^2} \quad (47)$$

Also from the Appendix

$$K = \frac{1}{\sigma_E\sqrt{(2\pi)}}\int_{-\infty}^{\infty} zf(z)\varepsilon^{-z^2/2\sigma_E^2}dz \quad . \quad . \quad . \quad . \quad (48)$$

$$\sigma_E^2 = \int_{-\infty}^{\infty} G_E(\omega)d\omega = \int_{-\infty}^{\infty} |Y_1(j\omega)|^2 G_e(\omega)d\omega \quad . \quad . \quad (49)$$

rom eqns. (47)–(49) K and σ_E can be eliminated and the elation between σ_e and σ_x obtained.

(5) THE STABILITY OF SECOND-ORDER NON-LINEAR CONTROL SYSTEMS WITH RANDOM INPUTS

For a control system of the type shown diagrammatically in ig. 5 the following equations hold:

$$\left.\begin{aligned} e &= x - y \\ E &= Y_1(p)e \\ Y_2(p)y &= f(E) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

vhich give

$$Y_2(p)e + fY_1(p)e = Y_2(p)x \quad . \quad . \quad . \quad . \quad (51)$$

f now $Y_2(p)$ is a second-order transfer function of the form

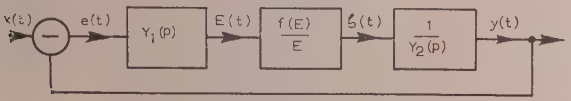


Fig. 5.—Second-order control system with random input.

$p^2 + Kp + L$, as it will be for a simple motor and load, and $_1(p)$ is first order, eqn. (49) will be of the form

$$\ddot{e} + k\dot{e} + le + f(e, \dot{e}) = \ddot{x} + k\dot{x} + lx \quad . \quad . \quad (52)$$

vhich can be rewritten

$$\ddot{e} + F(e, \dot{e}) = \ddot{x} + k\dot{x} + lx \quad . \quad . \quad . \quad . \quad (53)$$

3y one of various methods of construction the phase-plane rajectories can be drawn for the equation

$$\ddot{e} + F(e, \dot{e}) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

oy rewriting in the form

$$\frac{d\dot{e}}{de} = -\frac{F(e, \dot{e})}{\dot{e}} \text{ since } \ddot{e} = \dot{e}\frac{d\dot{e}}{de} \quad . \quad . \quad . \quad (55)$$

. It is now seen that if there is a finite input $x(t)$, the effect on he gradient of the trajectory is to increase it by

$$-\frac{\ddot{x} + k\dot{x} + lx}{\dot{e}}$$

ince from eqn. (51)

$$\frac{d\dot{e}}{de} = -\frac{F(e, \dot{e})}{\dot{e}} + \frac{\ddot{x} + k\dot{x} + lx}{\dot{e}} \quad . \quad . \quad . \quad (56)$$

t is reasonable to consider the input as made up of small steps f acceleration $\delta\ddot{x}$ and if the system is initially at some point P_0 e_0, $\dot{e}_0$), with input conditions $\ddot{x}_0$, $\dot{x}_0$, x_0, the representative oint will move in the direction P_0P_1 instead of along the trajectory. After a small interval of time δt, the representative oint will have reached a new trajectory at P_1 and will travel long a path making an angle with this new trajectory dependent n the new values of $\ddot{x}$, $\dot{x}$ and x. Now since random behaviour or $x(t)$ has been assumed, these values will be continually hanging, and the variations of actual gradient from that of the rajectories drawn for $x(t) = 0$ will be sometimes positive and ometimes negative, and the representative point will follow a lrunkard's course about one of the drawn trajectories. Now if he system is a sensible system for the type of input assumed, it is an essential condition that wherever the system starts initially after a sufficiently long time it will tend towards the origin ($e = 0$, $\dot{e} = 0$) of the phase-plane diagram. Thus, whatever the fluctuations the resulting trajectory must always have a bias towards the origin in Fig. 6. In any practcial system there must be upper and lower limits for the input $x(t)$, and so no matter how the input varies it must always perform some sort of drunkard's walk with reflecting boundaries at the upper and lower limits of $x(t)$; and therefore the greater the modulus of $\dot{x}$ and $\ddot{x}$ the shorter the time they can continue without changing sign. Thus, the greater the value of the correcting term

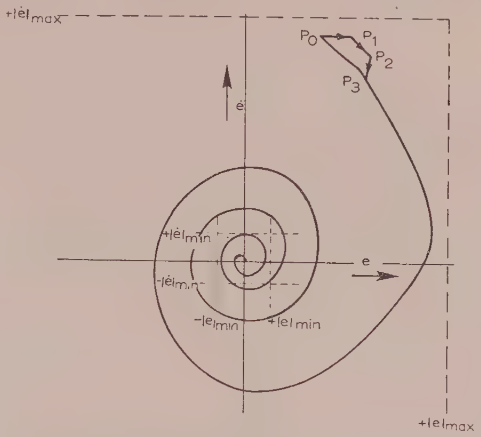


Fig. 6.—Phase-plane diagram of possible second-order system showing limits of error and error rate.

$$\frac{\ddot{x} + k\dot{x} + lx}{\dot{e}}$$

the shorter the time it can operate without changing sign. This sets a definite limit to the magnitude of the excursions from one of the trajectories drawn for $x(t) = 0$. Now not only will x, $\dot{x}$ and $\ddot{x}$ be bounded, but for physical reasons the output quantities y, $\dot{y}$ and $\ddot{y}$ will also be bounded, and under the very worst conditions therefore

$$|e|_{max} = |x|_{max} + |y|_{max},\ |\dot{e}|_{max} = |\dot{x}|_{max} + |\dot{y}|_{max}$$
$$\text{and } |\ddot{e}|_{max} = |\ddot{x}|_{max} + |\ddot{y}|_{max} \quad . \quad . \quad (57)$$

must also be bounded. In practice, owing to the nature of the system it may be possible to put lower limits on $|e|_{max}$ and $|\dot{e}|_{max}$ which can be shown on the diagram. Further, there will, in general, be some small value $|e|_{min}$ of e below which it does not matter how e varies, and this being so $\dot{e}$ can be just so large that the system will always run into alignment so that e does not exceed $|e|_{min}$ as shown in the diagram. In this way the required values of $|\dot{e}|_{min}$ are determined.

If now the phase-plane diagram for $x(t) = 0$ exhibits limit cycles, the limit cycles will be alternately stable and unstable, and it will be necessary to ensure that a stable limit cycle (or a stable point) is enclosed within the box made by

$$\pm\ |e|_{min} \text{ and } \pm\ |\dot{e}|_{min}$$

and that the box made by

$$\pm\ |e|_{max} \text{ and } \pm\ |\dot{e}|_{max}$$

is enclosed within the next limit cycle, which will be unstable. Then, provided that at some time the representative point starts within the e_{min} box, it will come out of it only when $\ddot{x}$ exceeds $\ddot{y}$

for a considerable period, $\ddot{y}$ being limited by the available output power. Provided that this extreme value of $\ddot{x}$ cannot operate for long enough to take the representative point outside the unstable limit cycle, the system must always tend to run back to the origin by a drunkard's course about one of the trajectories.

Reference to eqn. (54) shows that when $\dot{e}$ approaches zero a transformation in e can be made such that

$$F(e_x, 0) = \ddot{x} + k\dot{x} + lx \quad . \quad . \quad . \quad . \quad (58)$$

which is equivalent to a shift of the phase-plane diagram along the e-axis by an amount, which, for a system in which $l = 0$, will be the sum of the velocity and acceleration lags. This indicates that some adjustment of $|e|_{min}$ must be made to take into account these lags, and the stable limit cycle must, in fact, lie entirely within these reduced limits. Similar adjustments must be made to the values of $|e|_{max}$ and $|\dot{e}|_{max}$, and just how to assess these adjustments is one of the outstanding problems.

Thus, given the limits of $|e|_{min}$, $|\dot{e}|_{min}$, $|e|_{max}$ and $|\dot{e}|_{max}$, the problem in designing a system is to alter the parameters in such a way that the trajectories run by the shortest possible paths from any point within the outer box to some unspecified point on the boundary of the inner box. This implies the fewest possible crossings of the $\dot{e}$-axis, but it must be remembered that it is the time taken for the representative point to reach the inner box which is important rather than the actual distance. The presence or absence of limit cycles is then unimportant, provided that no part of one lies in the region between the inner and outer boxes.

With regard to the values of $|e|_{max}$ and $|\dot{e}|_{max}$, if the probability of large values can be calculated it will usually be found that the probability is very small indeed for the occurrence of values of e and $\dot{e}$ that are more than very small fractions of

$$|x|_{max} + |y|_{max} \text{ and } |\dot{x}|_{max} + |\dot{y}|_{max}$$

except under conditions of switching on. It may therefore be possible to arrange for some debasement of the system on switching on, e.g. by reducing the gain or increasing the damping, so that it runs into line more slowly and then automatically reverts to the high-performance state when the representative point lies within the dangerous limit cycle of the high-performance condition. It may then be expedient to arrange for the system to revert to the lower-performance state in the event of e and $\dot{e}$ attaining values which bring the representative point dangerously near the unstable limit cycle.

(6) CONCLUSIONS

Booton's approximation for the analysis of non-linear control systems with random inputs has been described and is justified in the Appendix. It has been found by numerical computation that, for a control system in which the motor and load is of second order and in which a phase advance network is included before a saturating amplifier, Booton's approximation gives excellent results. At present, however, no means are known of applying Booton's method to a non-linear system with a random input in the presence of noise when it is required to optimize the system so that the output follows the wanted input as closely as possible, the statistical nature of both wanted input and noise being known. Burt's method, on the other hand, can be used for this purpose, as has been shown, and it has been successfully employed in practice for a system containing a saturating element. Details of these results are not yet available for publication, but it is hoped that a paper from Mr. Burt will be available shortly.

The problem of how to determine performance criteria for non-linear systems still remains. From what has been said it appears essential to find criteria which are applicable to random inputs if the comparison of non-linear systems is to be meaningful. The mathematics developed for statistics is mostly based on the assumption that the variates involved are unbounded, and given a long enough time infinite values will be recorded. In physical control systems this is not the case, as both the input and output and their derivatives, and therefore the error function and all its derivatives, are bounded.

A possible approach which is being investigated is to consider a control system as transforming a bounded set of all possible messages at the input into a second bounded set of messages at the output, some suitably chosen norm of the difference being taken as the measure of performance. How this may be achieved has yet to be discovered and may, of course, be beyond the bounds of possibility.

(7) ACKNOWLEDGMENTS

The authors wish to acknowledge with thanks the encouragement and assistance of their colleagues in the Engineering Department of Cambridge University, with whom and with Messrs. C. A. A. Wass and E. G. C. Burt of the Ministry of Supply, they have had many useful discussions.

(8) REFERENCES

(1) Wiener, N.: "The Extrapolation, Interpolation and Smoothing of Stationary Time Series" (Wiley, New York, 1949).

(2) Lampard, D. G.: "The Response of Linear Networks to Suddenly Applied Stationary Random Inputs," *Journal of Applied Physics* (to be published).

(3) Rice, S. O.: "Mathematical Analysis of Random Noise," *Bell System Technical Journal*, 1944, **23**, p. 282 and 1945, **24**, p. 46.

(4) Rice, S. O.: *ibid.*, 1944, **23**, p. 35.

(5) Middleton, D.: "The Response of Biased Saturated Linear and Quadratic Rectifiers to Random Noise," *Journal of Applied Physics*, 1946, **17**, p. 778.

(6) Booton, R. C.: "Non-linear Control Systems with Statistical Inputs," Dynamic Analysis and Control Laboratory Report No. 61 (Massachusetts Institute of Technology, March, 1952).

(7) Courant, R., and Hilbert, D.: "Methods of Mathematical Physics" (Interscience, New York, 1953).

(8) Lampard, D. G.: "The Minimum Detectable Change in the Mean Noise-input Power to a Receiver," *Proceedings I.E.E.*, Monograph No. 80 R, November, 1953 (**100**, Part IV, p. 118).

(9) Lampard, D. G.: "Generalization of the Wiener–Khintchine Theorem to Non-stationary Processes," *Journal of Applied Physics*, 1954, **25**, p. 802.

(10) Booton, R. C., Mathews, M. V., and Seifert, W. W.: "Non-linear Servomechanisms with Random Inputs," Dynamic Analysis and Control Laboratory Report (Massachusetts Institute of Technology, August, 1953), p. 15).

(11) Seigert, A. F. J.: "The Passage of Stationary Processes through Linear and Non-linear Devices," Berkeley Symposium on Statistical Methods in Communication Engineering (Berkeley, California, August, 1953).

(12) Thomson, W. E.: "The Response of a Non-Linear System to Random Noise," *Proceedings I.E.E.*, Monograph No. 106 R, September, 1954 (**102** C, p. 46).

(9) APPENDIX

(9.1) Evaluation of the Integral $\int_{-\infty}^{\infty}\int f(x_1)f(x_2)p(x_1, x_2)dx_1dx_2$ where $p(x_1, x_2)$ is a two-dimensional symmetric probability density function.

The two methods mentioned in the text are

(*a*) Series expansion with possible approximation.
(*b*) Exact evaluation.

Exact evaluation is possible in only a few cases. The series expansion method is most conveniently used if p is a function only of the variables, x_1/σ, x_2/σ and ρ, where σ is the standard deviation of x_1 and x_2 and ρ is the autocorrelation function (normalized).

(9.1.1) Series Expansion.

This method is very convenient when $p(x_1, x_2)$ is a two-dimensional Gaussian distribution, since Mehler's expansion (see below) may be used. Analogous expansions exist also for other probability distributions.*

(9.1.1.1) *The Gaussian Case.*

Here

$$p(x_1, x_2) = \frac{1}{2\pi\sigma^2\sqrt{(1-\rho^2)}}\exp\left[\frac{-(x_1^2+x_2^2-2\rho x_1x_2)}{2\sigma^2(1-\rho^2)}\right]$$

$$= \frac{1}{2\pi\sigma^2\sqrt{(1-\rho^2)}}\exp\left[\frac{-(\xi_1^2+\xi_2^2-2\rho\xi_1\xi_2)}{2(1-\rho^2)}\right]$$

where $\xi_1 = x_1/\sigma,\ \xi_2 = x_2/\sigma$ . . . . . (59)

Theorem (Mehler's Expansion).

$$\frac{1}{\sqrt{[\pi(1-\rho^2)]}}\exp\left(-\frac{x^2+y^2-2\rho xy}{1-\rho^2}\right)$$

$$= \exp-(x^2+y^2)\sum\frac{H_n(x)H_n(y)}{2^nn!\sqrt{\pi}}\rho^n \quad . \quad (60)$$

where $H_n(x)$, the nth Hermite polynomial $= (-1)^n\varepsilon^{x^2}d^n/dx^n\varepsilon^{-x^2}$. The result may be shown by the following straightforward method or, equivalently, by a two-dimensional Fourier transformation of both sides:

$$\varepsilon^{-x^2} = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\exp(-u^2+2jxu)du \quad . \quad . \quad (61)$$

Therefore $H_n(x) = (-1)^n\varepsilon^{x^2}\dfrac{d^n}{dx^n}\varepsilon^{-x^2}$

$$= \frac{(-2j)^n\varepsilon^{x^2}}{\sqrt{\pi}}\int_{-\infty}^{\infty}u^n\exp(-u^2+2jxu)du \quad . \quad (62)$$

Therefore $\displaystyle\sum_{n=0}^{\infty}\frac{H_n(x)H_n(y)\rho^n}{2^nn!\sqrt{\pi}} = \sum_{n=0}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{(2j)^{2n}\rho^nu^nv^n}{2^nn!\pi^{3/2}}\varepsilon^{x^2+y^2}$

$$\times\exp(-u^2-v^2+2jxu+2jyv)dudv \quad . \quad (63)$$

$$|\rho| \leqslant \rho_1 < 1$$

but $\displaystyle\sum_{n=0}^{\infty}\frac{(-2\rho uv)^n}{n!} = \varepsilon^{-2\rho uv}$

Therefore r.h.s. $= \dfrac{\exp(x^2+y^2)}{\pi^{2/3}}\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(-u^2-2\rho uv$

$$-v^2+2jxu+2jyv)dudv$$

using (61), $\displaystyle\int_{\infty}^{\infty}\exp(-v^2-2\rho uv+2jyv) = \sqrt{\pi}\exp[-(y+ju\rho)^2]$

Therefore r.h.s. $= \dfrac{\varepsilon^x}{\pi}\displaystyle\int_{-\infty}^{\infty}\exp[-(1-\rho^2)u^2+2j(x-\rho y)u]du$

and again using eqn. (61)

$$\text{r.h.s.} = \frac{\exp(x^2+y^2)}{\sqrt{[\pi(1-\rho^2)]}}\exp\left(-\frac{x^2+y^2-2\rho xy}{1-\rho^2}\right) \quad . \quad (64)$$

Mehler's expansion now follows on multiplying both sides by

$$\exp-(x^2+y^2)$$

In probability applications it is convenient to use a form of the theorem based on $\varepsilon^{-\xi^2/2}$. This follows by putting

$$x = \frac{\xi_1}{\sqrt{2}} \qquad y = \frac{\xi_2}{\sqrt{2}}$$

and introducing the Hermite polynomials $X_n(\xi)$ based on $\varepsilon^{-\xi^2/2}$, namely

$$X_n(\xi) = \varepsilon^{\xi^2/2}\frac{(-1)^n}{\sqrt{(n!)}}\frac{d^n}{d\xi^n}\varepsilon^{-\xi^2/2} \quad . \quad . \quad . \quad . \quad (65)$$

The first few functions X_n are

$$X_0(\xi) = 1$$
$$X_1(\xi) = \xi$$
$$X_2(\xi) = \frac{1}{\sqrt{2}}(1-\xi^2)$$
$$X_3(\xi) = \frac{1}{\sqrt{\sigma}}(\xi^3-3\xi)$$

Mercer's formula now becomes

$$\frac{1}{2\pi\sqrt{(1-\rho^2)}}\exp\left[\frac{-\xi_1^2+\xi_2^2-2\rho\xi_1\xi_2}{2(1-\rho^2)}\right]$$

$$= \frac{\varepsilon^{-(\xi_1^2+\xi_2^2)/2}}{2\pi}\sum_{n=0}^{\infty}X_n(\xi_1)X_n(\xi_2)\rho^n \quad . \quad (66)$$

Changing the variables in the required integral to ξ_1 and ξ_2 and putting in this value

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x_1)f(x_2)p(x_1, x_2)dx_1dx_2$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(\sigma\xi_1)f(\sigma\xi_2)\varepsilon^{-(\xi_1^2+\xi_2^2)/2}\sum_{n=0}^{\infty}\rho^nX_n(\xi_1)X_n(\xi_2)d\xi_1d\xi_2$$

$$= \frac{1}{2\pi}\sum_{n=0}^{\infty}\rho^n\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(\sigma\xi_1)f(\sigma\xi_2)\varepsilon^{-(\xi_1^2+\xi_2^2)/2}X_n(\xi_1)X_n(\xi_2)d\xi_1d\xi_2$$

$$= \sum_{n=0}^{\infty}a_n^2\rho^n \quad . \quad . \quad . \quad . \quad . \quad . \quad (67)^*$$

where $a_n = \dfrac{1}{\sqrt{(2\pi)}}\displaystyle\int_{-\infty}^{\infty}f(\sigma\xi)X_n(\xi)\varepsilon^{-\xi^2/2}d\xi$ since the double integrals split up into the product of two identical single integrals. a_n is the coefficient of $X_n(\xi)$ in the expansion of $f(\sigma\xi)$ in terms of these polynomials. If f is an odd function, i.e. $f(-x) = -f(x)$ the terms, a_n, vanish when n is odd, since X_n is odd or even with n.

* All these seem to be Mercer expansions of the kernel $P(x_1, x_2, \rho)/\sqrt{[P(x_1)P(x_2)]}$. See Reference 7, pp. 134 *et seq.*

* Since the paper was written a similar method of obtaining the autocorrelation function of the output of a non-linearity in terms of that of the input has been published by Thomson.[12]

Connection with Booton's[6] *Approximation.*

$$a_1 = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} f(\sigma\xi)\xi\varepsilon^{-\xi^2/2}d\xi = \sigma K \quad . \quad . \quad . \quad (68)$$

where K is Booton's constant which he called the equivalent gain of the element. Thus for a symmetrical device (f an odd function) the expansion of the output autocorrelation function is

$$\left.\begin{aligned}\phi_2 &= (\sigma K)^2\rho + \text{terms in } \rho \text{ of degree} \geqslant 3\\ &= K^2\phi_1 + \ldots\end{aligned}\right\} \quad . \quad . \quad (69)$$

where ϕ_1 and ϕ_2 are the input and output autocorrelation functions. Consequently Booton's approximation, which would be equivalent to retaining only the first term of this expansion, is justified when the remaining terms are negligible. This will be the case when the non-linear element is followed by a low-pass filter with bandwidth which is small compared with that of the error signal, as shown below.

Suppose that the output signal of the non-linear device is passed through a filter of impulse response $h(t)$. Let the autocorrelation function of the output of this filter be ϕ_3. This is related to the input autocorrelation function by

$$\phi_3(\tau) = \int_{-\infty}^{\infty}\phi_2(\tau')H(\tau-\tau')d\tau'$$

where $$H(\tau') = \int_{-\infty}^{\infty} h(\tau'+\tau'')h(\tau'')d\tau'' \quad . \quad . \quad . \quad (70)$$

is what Lampard has called the filter autocorrelation function. It has some properties of an autocorrelation function

(i) $H(0) \geqslant H(\tau)$ all τ

(ii) $H(\tau) \to 0$ as $\tau \to \infty$

Substituting

$$\phi_2(\tau) = \sum_{n=0}^{\infty} a_n^2\rho^n(\tau) \text{ the result is } \phi_3(\tau) = \int_{-\infty}^{\infty}\Sigma\, a_n^2\rho^n(\tau')H(\tau-\tau')d\tau'$$

$$= \sum_{n=0}^{\infty} a_n^2\int_{-\infty}^{\infty}\rho^n(\tau')H(\tau-\tau')d\tau' \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (71)$$

The change of summation and integration is justified by absolute convergence of $\Sigma a_n^2\rho^n$.

Now assuming $f(x)$ is odd, all terms in the summation corresponding to even n will vanish and the error in assuming a first-term approximation will be

$$\sum_{n=3,5,7,\ldots} a_n^2\int_{-\infty}^{\infty}\rho^n(\tau)H(\tau-\tau')d\tau' \quad . \quad . \quad . \quad (72)$$

In absolute value this is less than or equal to

$$\sum_{n=3,5,7} a_n^2\int_{-\infty}^{\infty}|\rho^n(\tau)|\,|H(\tau-\tau')|d\tau$$

$$\leqslant H(0)\sum_{n=3,5,7} a_n^2\int_{-\infty}^{\infty}|\rho^n(\tau)|d\tau$$

$$\leqslant H(0)\sum_{n=3,5,7,\ldots} a_n^2\int_{-\infty}^{\infty}\rho^2(\tau)d\tau$$

$$= \int_{-\infty}^{\infty}h^2(\tau)d\tau\int_{-\infty}^{\infty}\rho^2(\tau)d\tau(\sigma_2^2 - K^2\sigma^2) \quad . \quad (73)$$

where $\sigma_2^2 = \phi_2(0)$, and gives a rough upper limit for the error in $\phi_3(\tau)$. In practice, the error may be expected to be much smaller as $\rho^n(\tau)$ tends to a function $g(\tau)$ where $g(0) = 1, g(\tau) = 0$, $\tau \neq 0$.

(9.1.1.2) *Extensions of the Above Method.*

Analogous expansions exist also for other distributions, e.g. square-law envelope (Lampard).[8]

An extension of the method can be used to give the higher output moments $\int \ldots \int f(x_1)f(x_2)\ldots f(x_n)\text{p}(x_1\ldots x_n)dx_1\ldots dx_n$— in the case of a Markoff process.

Lampard[2] has also given an extension of the method for calculating moments for non-stationary series.

(9.1.2) **Explicit Evaluation of** $\int_{-\infty}^{\infty} f(x_1)f(x_2)p(x_1x_2)dx_1dx_2$.

The integral has been explicitly evaluated, i.e. in terms of tabulated functions, in only a few cases. If $f(x)$ consists of straight segments, the integral may be expressed in terms of the one-dimensional distribution function—in the Gaussian case an error function.

Example (a).

On-off with Gaussian input

$$f(x) = \begin{Bmatrix} +x_0 & x > 0\\ -x_0 & x < 0\end{Bmatrix} \quad . \quad . \quad . \quad (74)$$

$$f(0) = 0$$

Since f is odd, the integral splits up as follows:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} = \int_0^{\infty}\int_0^{\infty} + \int_0^{\infty}\int_{-\infty}^{0} + \int_{-\infty}^{0}\int_{-\infty}^{0} + \int_{-\infty}^{0}\int_0^{\infty} = 2\int_0^{\infty}\int_0^{\infty} - 2\int_0^{\infty}\int_0^{\infty}$$

$$(\rho \text{ replaced by } -\rho) \quad . \quad (75)$$

$$= \frac{2x_0^2}{2\pi\sigma^2\sqrt{(1-\rho^2)}}\left[\int_0^{\infty}\int \exp -(\xi_1^2+\xi_2^2-2\rho\xi_1\xi_2)d\xi_1d\xi_2\right.$$

$$\left.-\int_0^{\infty}\int \exp -(\xi_1^2+\xi_2^2+2\rho\xi_1\xi_2)d\xi_1d\xi_2\right]2\sigma^2\sqrt{(1-\rho^2)}$$

$$= \frac{2x_0^2}{\pi}\left[\frac{(\pi-\psi)-\psi}{2}\right] \text{ where } \rho = \cos\psi \text{ (see below)}$$

$$= \frac{2x_0^2}{\pi}\left(\frac{\pi}{2}-\psi\right) = x_0^2\left(\frac{\text{arc sin }\rho}{\pi/2}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad (76)$$

It remains to justify the last step but one. If $\psi \neq 0,\pi$,

$$\int_0^{\infty}\int_0^{\infty}\varepsilon^{-(\xi_1^2+\xi_2^2+2\xi_1\xi_2\cos\psi)}d\xi_1d\xi_2 = \text{cosec }\psi\int_0^{\infty}\int_0^{\infty}\varepsilon^{-(\eta_1^2+\eta_2^2)}d\eta_1d\eta_2$$

$$. \quad . \quad . \quad . \quad (77)$$

where $$\eta_1 = \cos\frac{\psi}{2}(\xi_1+\xi_2),\ \eta_2 = \sin\frac{\psi}{2}(-\xi_1+\xi_2)$$

Now put $$r^2 = \eta_1^2 + \eta_2^2,\ \theta = \text{arc tan}\,\frac{\eta_2}{\eta_1}$$

$$\int_0^\infty\int_0^\infty \varepsilon^{-(\eta_1^2+\eta_2^2}d\eta_1 d\eta_2 = \int_{-\psi/2}^{\psi/2}\int_0^\infty \varepsilon^{-r^2} r\,d\theta\,dr = \frac{\psi}{2} \quad . \quad (78)$$

Similarly, by changing ψ to $\pi - \psi$

$$\int_0^\infty\int \varepsilon^{(\xi_1^2+\xi_2^2-2\xi_1\xi_2\cos\psi)}d\xi_1 d\xi_2 = \frac{\pi-\psi}{2}$$

Example (*b*).

If $f(x)$ consists of a finite number of straight-line segments (Fig. 7) the integral reduces to a sum of integrals of the types

$$\left.\begin{aligned}&\int_{a_1}^{b_1}\int_{a_2}^{b_2} P(\xi_1, \xi_2)d\xi_1 d\xi_2\\ &\int_{a_1}^{b_1}\int_{a_2}^{b_2} \xi_1 P(\xi_1\xi_2)d\xi_1 d\xi_2\\ &\int_{a_1}^{b_1}\int_{a_2}^{b_2} \xi_2 P(\xi_1\xi_2)d\xi_1 d\xi_2\\ &\int_{a_1}^{b_1}\int_{a_2}^{b_2} \xi_1\xi_2 P(\xi_1\xi_2)d\xi_1 d\xi_2\end{aligned}\right\} \quad . \quad . \quad . \quad (79)$$

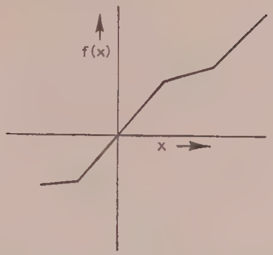


Fig. 7.—Function consisting of a number of straight-line segments.

In the Gaussian case, the following integrals are useful:[4]

$$\int_0^\infty\int_0^\infty \varepsilon^{-\xi_1^2-\xi_2^2-2\xi_1\xi_2\cos\psi}d\xi_1 d\xi_2 = \tfrac{1}{2}\psi \text{ cosec } \psi \quad . \quad (80)$$

$$\int_0^\infty\int_0^\infty \xi_1\xi_2\varepsilon^{-\xi_1^2-\xi_2^2-2\xi_1\xi_2\cos\psi}d\xi_1 d\xi_2 = \tfrac{1}{4}\text{cosec}^2\,\psi(1 - \psi\cot\psi) \ . \ (81)$$

$$\int_0^\infty\int_0^\infty \xi_i\varepsilon^{-\xi_1^2-\xi_2^2-2\xi_1\xi_2\cos\psi}d\xi_1 d\xi_2 = \frac{\sqrt{\pi}}{4}\,\frac{1}{1+\cos\psi} \quad . \quad (82)$$

for $i = 1, 2$.

621.391

Monograph No. 155 R
Nov. 1955

SOME TERMINOLOGY AND NOTATION IN INFORMATION THEORY

By I. J. GOOD, M.A., Ph.D.

(*The paper was first received* 18*th July, and in revised form* 10*th September*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in November*, 1955.)

SUMMARY

The main purpose of the paper is to stimulate thought concerning terminology, notation and exposition of some basic parts of information theory. The notation used here is intended to be simple, nearly self-explanatory, readable in words from left to right, and suggestive of new applications. Moreover, sufficient generality is preserved to ensure that entropy can be interpreted without necessarily depending on the frequentist definition of probability (as a limiting frequency in an infinite sequence of trials). Mention is made of some connections of the theory with inverse probability and with mathematical statistics in general.

(1) PREFERENCE FOR THE TERMINOLOGY OF THE MORE FUNDAMENTAL SCIENCES

A serious difficulty in the unification of science is that different branches use different words for the same thing and the same word for different things. A principle that would help to overcome this difficulty is that the words (and notations) used in the more fundamental branches of science should generally be preferred when there is a choice. For example, the word "average" would be better replaced by "expectation" (or "expected value") when that is what it means. It is also desirable to define technical words in terms of fundamental sciences: an ensemble could be defined as a population of infinite sequences of trials, and in fact could in most contexts be called a population. If an instructor does not wish to run counter to established usage he can at least point out where established usage is unsatisfactory. For example, the word "expectation" in the theory of probability is unsatisfactory, but once this has been emphasized it no longer leads to confusion.

It is too much to hope that engineers will be prepared to accept the terminology of probability and statistics, but perhaps some engineers may be interested to see how information theory can be expressed in such terminology. If two languages must be used it may be helpful to show that translation is possible.

Within the theory of probability itself there is further room for improvement. When applying Bayes's theorem, the expressions *a priori* and *a posteriori* are unnecessarily pompous and have, as Jeffreys[1] has pointed out, misleading philosophical associations. The only justification for them is that they are "O.K. words" in the sense of Potter.[2] Jeffreys shortened them to prior and posterior and at a recent meeting of a philosophy of science group he used the even better terms initial and final, previously suggested by von Mises.[3] These terms have the further advantage that one can interpolate intermediate probabilities in sequential problems: the Latin for intermediate is not known to every schoolboy.[4]

(2) PROBABILITY NOTATION

One of the main apparent dichotomies of notation in the theory of probability is concerned with probabilities of propositions on the one hand and probabilities associated with random variables on the other. Although probabilities associated with random variables are of more direct interest to most engineers, the more fundamental idea is that of the probability of a proposition, and all other types of probability are on that basis special cases, provided that the word proposition is taken in its most general sense.[*] All that is required of a proposition is that there should be a meaning in saying that it is true or that it is false;[8] the meaning of meaning will not be discussed here. Propositions are part of the stock-in-trade of logic, a subject that is regarded as more fundamental even than probability. We shall therefore regard the notation $P(F|G)$, where F and G are propositions, as the fundamental notation in probability theory. This notation is used, for example, by Jeffreys[1] and is read "the probability of F given G." Probabilities themselves may be interpreted in the most general sense, i.e. in the subjectivistic sense, as in References 7 and 9. The proposition G can sometimes be taken for granted and omitted from the notation without confusion, although its omission sometimes does cause confusion.

Hypotheses will be regarded as a special type of proposition. If H is a hypothesis and E is the proposition describing an event or experimental result, then $P(E|H)$ is a more familiar expression than $P(H|E)$; but both types of expression will be permitted in the paper, because $P(F|G)$ is regarded as existing for every pair of propositions. Only the six symbols E, F, G, H, K, L (possibly with suffixes) will be used for propositions, and E and H when it is required to suggest the intuitive interpretation of propositions as experimental results (or events) and as hypotheses. Any such suggestion will be made only because of the general interest in experiments and hypotheses, and it will always be legitimate to use F and G, for example, instead of E and H, and to interpret F and G as quite general propositions.

The notation $P(x)$, meaning the probability that a random variable, X, takes the value x, will be used in the later Sections. In the propositional notation, $P(x)$ should be written $P(X = x)$, and $P(x)$ is best regarded as an abbreviated notation.

For continuous variables with probability densities, there is something to be said for the notation $p(x)$, meaning

$$\lim_{\delta x \to 0} \frac{P(x < X < x + \delta x)}{\delta x}$$

but other notations, such as $P(dx)$ or $P(x)$, also occur in the literature.[†] In the present paper discrete variables only will be used: this reference to continuous variables is made in order to show that probability has its need of accepted conventions just as much as information theory.

(3) AMOUNT OF INFORMATION IN A PROPOSITION[‡]

Continuing with the fundamental, i.e. simple-minded, approach, we may define the amount of information in a proposition F as minus the logarithm of its probability, $-\log P(F)$, or as $-\log P(F|G)$ when some hypothesis, G, is regarded as given.[6,7] It is only when we come to the notion of entropy (defined in

[*] The use of random variables has the same generality as that of propositions, but only in an artificial manner.

[†] Jeffreys[1] (2nd edition, p. 120) gives reasons for the notation $P(dx)$, meaning $p(x)dx$.

[‡] Not to be confused with the amount of information, available in a sample, concerning a population parameter, as defined by Fisher.[5]

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Good is in the Royal Naval Scientific Service.

Section 5) that expectations will be taken. We may write $I(F) = -\log P(F)$.

The information may be regarded as the amount received, concerning F, when we discover that F is true, i.e. when its probability becomes unity in virtue of new evidence.

It is sometimes thought that the measure of the amount of information in a proposition should depend on the utility of the proposition if true. We might, for example, say that the amount of information in the Proceedings of the Second London Symposium on Information Theory was 65 shillings'-worth. To exclude utilities from the definition of amount of information depends simply on a decision. If we decide to make the amount of information depend only on probabilities and if we insist on its having the additive property for probabilistically independent propositions, then minus the logarithm of the probability is the only possible definition. The base of the logarithms determines the unit of measurement, while the minus sign is inserted merely in order to make the measure non-negative when the base is greater than unity. An amount of information may more specifically be called an amount of probabilistic information. Extensions to "utilitarian information" would also be of interest.[15,22]

(4) AMOUNT OF INFORMATION IN ONE PROPOSITION CONCERNING ANOTHER

Denote by E the proposition stating the result of an observation or experiment (the evidence). Suppose that, when E is given, the probability of H does not become unity, i.e. that $P(H|E) \neq 1$. Then the extra amount of information that would be obtained if H were now found to be true would be $-\log P(H|E)$. Hence it is natural to say that the amount of information concerning H provided by E is

$$-\log P(H) - [-\log P(H|E)] \text{ units,}$$

i.e.

$$\log \frac{P(H|E)}{P(H)} \text{ units,}$$

i.e.

$$\log \frac{P(E \text{ and } H)}{P(E)P(H)} \text{ units}$$

if E has non-zero probability, a condition that will always be satisfied in practice. The last expression is the logarithm of what may be called the association factor between E and H. It is symmetrical in E and H. This natural generalization of the above definition of the amount of information in a proposition is used by Woodward[10] and Goldman,[11] but with different terminology from that used here and with less apparent desire for philosophical generality. We may say that

$$\log \frac{P(F \text{ and } G)}{P(F)P(G)} = I(F:G), \text{ say}$$

where F and G are arbitrary propositions of non-zero probabilities, is the amount of information in G concerning F, and is also the amount in F concerning G, in virtue of the symmetry of the expression. If F and G are probabilistically independent, each gives no information concerning the other, since then $P(F \text{ and } G) = P(F)P(G)$. Another special case is when $F = G$, the consideration of which is left as an exercise for the reader.

A very simple example of a communication or information system occurs when someone tells you that H is true. If he is reliable, and is known to be reliable, then the amount of information received concerning H is $-\log P(H)$, but if he is unreliable or dishonest then it is only $-\log P(H) + \log P(H|E)$, which is less than $-\log P(H)$, where E is the proposition or event that the man has told you that H is true. An unreliable or dishonest man is an example of a noisy transducer. Moreover liars and writers of bad prose change the statistics of language itself and therefore make it more difficult for their betters to convey information.

The definition of $I(F:G)$ connects up with that of a sufficient statistic. If E is an experimental result and θ is a statistic, i.e. a numerical or vector function of E, then $\hat{\theta}$ is said to be sufficient for θ (a population parameter) if $P(E|\hat{\theta}) = P(E|\theta \text{ and } \hat{\theta})$. (The values of other population parameters may be taken for granted throughout.) It can be deduced that $I(\theta : E) = I(\theta : \hat{\theta})$, and this gives rigorous meaning to the familiar statement that $\hat{\theta}$ provides all the information concerning θ that is provided by the experiment or evidence. This connection between information theory and sufficient statistics is simpler than one that has been suggested previously, using expected amounts of information.[26]

An application of the association log-factor has been made to the theory of contingency tables, especially for the estimation of the probabilities of events that have never happened before.[12] Contrary to a crude definition of probability that is sometimes given, these probabilities are not zero. This work[12] makes use of the after-effect function originally tabulated for an electrodynamical application.[13]

It is of some historical and logical interest to note how inverse probability is related to the amount of information in a proposition concerning another one.

Let $H_i \, (i = 1, 2, 3, \ldots)$ be a sequence of possible hypotheses, and E a proposition describing an event. By Bayes's theorem, which is itself an immediate consequence of the product axiom in the theory of probability, $P(H_i|E)/P(H_i) = P(E|H_i)/P(E)$, and this is proportional to the "likelihood" $P(E|H_i)$, when E is fixed. Thus the amounts of information in E concerning the H's form a set of relative log-likelihoods, also called "relative log-factors" or "relative weights of evidence" in favour of the various hypotheses.[7] The word relative is used here to mean that any constant can be added to all the expressions, not that they can all be multiplied by a constant, without a change in the base of the logarithms.

A further relationship between information theory and inverse probability is of interest in the problem of fair fees to be paid to experts who estimate probabilities, such as meteorologists, technical advisers to firms, and other tipsters.[14,15] A still further relationship is mentioned below.

(5) ENTROPY

Suppose we have a number of mutually exclusive and exhaustive hypotheses $H_1, H_2, H_3, \ldots$ In some communication systems these hypotheses correspond to the letters of the alphabet or of a generalized alphabet, sometimes called a "sample space." Suppose that we intend to make an observation that will decide definitely which of the H's is true. If H_i is found to be true, the amount of information received concerning H_i, i.e. concerning the truth about the H's, is $-\log P(H_i)$. Therefore, before the observation is made, the expected amount of information concerning the H's, to be obtained from the observation, is $-\sum_i P(H_i) \log P(H_i)$, usually called the entropy.

More fully it may be called the entropy of the H's provided by the method of observation. It seems desirable that it should never be called simply an amount of information but rather an expected amount. In a long series of trials that are essentially the same and are probabilistically independent, the average amount of information received per trial will be approximately equal to the expected amount. This statement follows from the law of large numbers in the theory of probability. The word "average" is intended here in its usual arithmetical sense. Since the average in a finite series of trials (and all actual series are in fact finite) is only approximately equal to the expectation, it is a little misleading to confuse the average and the expectation, as has often been done in the literature both of information theory and of theoretical physics.

The approximate equality of the average information per trial and the expected amount per trial extends also to Markovian sequences of trials, where the influence of the past extends only over a bounded time.

Suppose, now, that apart from the hypotheses H_1, H_2, $H_3, \ldots$ there are a number of mutually exclusive and exhaustive experimental results E_1, E_2, $E_3, \ldots$ These are not intended to represent the results of a sequence of trials, but simply the possible results of a single trial. In communication theory, which is a branch of information theory, the H's are hypotheses describing, without reference to the meanings, what were the symbols or messages sent, and the E's are the propositions or hypotheses for the possible symbols or sequences of symbols received. The amount of information in E_j concerning H_i is

$$\log \frac{P(H_i \text{ and } E_j)}{P(H_i)P(E_j)}$$

and if H_i is true (and perhaps later discovered to be true) we may describe this expression as the amount of information in E_j concerning the truth about the H's. If E_j is observed, the probability of H_i is $P(H_i|E_j)$; hence the expected amount of information in E_j concerning the H's is

$$\sum_i P(H_i|E_j) \log \frac{P(H_i \text{ and } E_j)}{P(H_i)P(E_j)}$$

This is a function of j but not of i. It may be denoted by ent $(\mathbf{H}:E_j)$, where $\mathbf{H}$ is the class of H's. We may read this expression as "the entropy concerning the H's provided by E_j." The simpler form of entropy previously mentioned may be written ent $\mathbf{H}$. Those who do not like the word entropy may prefer to replace ent by $\mathscr{E}I$, read "expected amount of information concerning"; Bar-Hillel is understood to advocate this notation.

We have used a colon in ent $(\mathbf{H}:E_j)$, instead of a vertical stroke for the following reason: In the probability notation $P(G)$ there is almost always some proposition, K, often very complicated, that has been taken for granted and omitted from the notation, i.e. $P(G)$ is an abbreviation for $P(G|K)$. Similarly $P(G|L)$ is often an abbreviation for, say, $P(G|K \text{ and } L)$. Thus $-\sum_i P(H_i)$ usually means $-\sum_i P(H_i|K) \log P(H_i|K)$, for some K, and if we wish to bring K into the notation we are virtually forced to write ent $(\mathbf{H}|K)$, meaning "the entropy of the H's (or of H given K." So we must distinguish between "entropy in . . ." or "entropy provided by . . ." on the one hand and "entropy given . . ." on the other. Notation like ent $(\mathbf{H}\colon E_j|K)$, where K represents a single proposition since it is not in bold-face type, is now self-explanatory.

The expectation of ent $(\mathbf{H}:E_j)$, before E_j is observed, is

$$\sum_j P(E_j) \text{ ent } (\mathbf{H}:E_j)$$

which is a function of neither i nor j, and may be denoted by ent $(\mathbf{H}:\mathbf{E})$. It may be described as the entropy concerning the H's, or concerning the matter under investigation, provided by the E's, or provided by the method of observation, or provided by the experimental design.* It is equal to

$$\sum_{i,j} P(H_i \text{ and } E_j) \log \frac{P(H_i \text{ and } E_j)}{P(H_i)P(E_j)} = \text{ent } (\mathbf{H}) + \text{ent } (\mathbf{E}) - \text{ent } (\mathbf{H} \text{ and } \mathbf{E})$$

and by symmetry it is equal to ent $(\mathbf{E}:\mathbf{H})$.

If we now imagine the whole experiment to be one of a long series of independent trials, then the average amount of information concerning the H's is approximately equal to the conditional entropy provided by the method of observation. This statement naturally has its Markovian generalization.

(6) ENTROPY AND INVERSE PROBABILITY

Suppose that the number of possible experimental results $E_1, E_2, \ldots, E_n$ is finite and let H_0 be the equiprobable hypothesis, which makes $P(E_j|H_0)$ independent of j and therefore equal to $1/n$. Then H_0 maximizes ent $(\mathbf{E}|H)$ and in fact makes it equal to $\log n$. Let H_1 be some other hypothesis. Consider a sequence of independent trials each of whose results can be E_1, or E_2, or E_3, . . ., and consider how we may evaluate the evidence that H_1 is true rather than H_0. When E_j occurs, H_1 receives a factor, on its odds,* of $nP(E_j|H)$. The expected log-factor in favour of H_1, per trial, when H_1 is true, is

$$\sum_j P(E_j|H_1) \log [nP(E_j|H_1)] = \log n - \text{ent } (\mathbf{E}|H_1)$$
$$= \text{ent } (\mathbf{E}|H_0) - \text{ent } (\mathbf{E}|H_1)$$

This remark establishes a further connection between information theory and inverse probability. At the same time it forces expressions like $\sum_i p_i(\log p_i)^s$ on one's attention, when investigating the variance and higher moments of the log-factor or of the amount of information.[17] In fact, every generalized moment of the form $\sum_i p_i^r(\log p_i)^s$ may be regarded as a measure of heterogeneity of a multinomial distribution for which the chances of the various categories are $p_1, p_2, p_3, \ldots$[18]

If our two hypotheses are both general, i.e. neither is H_0, the above discussion can be generalized by the use of "cross-entropy," i.e. an expression of the form

$$\sum_j P(E_j|H_1) \log P(E_j|H_2)$$

The significance of cross-entropy in statistical mechanics had been foreseen by the author[19] and has since been confirmed by Professor B. O. Koopman. On the other hand Bartlett has pointed out that the last sentence on page 75 of Reference 7 is incorrect.

(7) THE RANDOM-VARIABLE NOTATION

So far, our notation has been propositional. The formulae become a little shorter if expressed in terms of random variables. Shannon[20] uses x and i corresponding to $\mathbf{H}$ and H_i in the above discussion; more precisely his x is a random variable that takes values i. In the theory of probability it is quite customary to denote random variables by capital letters. We may accordingly write X for Shannon's x, and x for his i, Y for his y and y for his j. Corresponding to ent $(\mathbf{H}:\mathbf{E})$, ent $(\mathbf{H}:E_j)$ and ent $(\mathbf{H}|E_j)$ we write ent $(X:Y)$, ent $(X:y)$ and ent $(X|y)$, the last two entropies, but not the first, being functions of y. As an example of the random-variable notation we have

$$\begin{aligned} \text{ent } (X:Y) &= \text{ent } (Y:X) \\ &= \mathscr{E}_{x,y} \log \frac{P(x \text{ and } y)}{P(x)P(y)} \\ &= \sum_{x,y} P(x \text{ and } y) \log \frac{P(x \text{ and } y)}{P(x)P(y)} \end{aligned}$$

* H and E may be regarded either as classes or as random propositions that take values H_i and E_j. They can be taken as the abstract definitions of "the matter under investigation" and "the experimental design."

* The odds corresponding to a probability p are defined as the number $p/(1-p)$, and the Bayes factor in favour of a hypothesis H is the number by which its initial odds are to be multiplied to get its final odds. When there are only two simple statistical hypotheses, this factor is equal to the likelihood ratio,[7,16] which is here $P(E_j|H_1)/P(E_j|H_0)$. The log-factor is also called the "weight of evidence in favour of H_1" or the "support for H_1."

$$
\begin{aligned}
&= -\sum_{x} P(x)\log P(x) - \sum_{y} P(y)\log P(y) \\
&\qquad + \sum_{x,y} P(x \text{ and } y)\log P(x \text{ and } y) \\
&= \text{ent } X + \text{ent } Y - \text{ent } (X \text{ and } Y) \\
&= \text{ent } X - \underset{y}{\mathscr{E}}\, \text{ent } (X|y) \\
&= \text{ent } Y - \underset{x}{\mathscr{E}}\, \text{ent } (Y|x)
\end{aligned}
$$

nce

$$
\begin{aligned}
\underset{y}{\mathscr{E}}\, \text{ent } (X|y) &= -\underset{y}{\mathscr{E}}\,\underset{x}{\mathscr{E}} \log P(x|y) \\
&= -\sum_{x,y} P(x \text{ and } y)\log \frac{P(x \text{ and } y)}{P(y)} \\
&= \text{ent } (X \text{ and } Y) - \text{ent } Y
\end{aligned}
$$

We may write ent $(X|Y)$ for $\underset{y}{\mathscr{E}}$ ent $(X|y)$, not to be confused ith ent $(X{:}Y)$. As before, the vertical stroke may be read given" and the colon may indicate "provided by" or "in." he use of the capital letter for a random variable (or Clarendon ype in the propositional notation) means that expectations have een taken with respect to that variable. The notation of colons nd vertical strokes may be used, in a sense that is now selfxplanatory, for the "unexpectated" information I, and it can e easily proved that

$$I(x \text{ and } y|z) = I(x|z) + I(y|x \text{ and } z)$$

$$I(x : y \text{ and } z) = I(x : y) + I(x : z|y),$$

oth of which relationships have a clear intuitive meaning, specially when read in words. By taking expectations in arious ways we may obtain identities concerning entropies. or example,

$$\text{ent } (X : Y \text{ and } Z) = \text{ent } (X : Y) + \text{ent } (X : Z|Y).$$

he special case $Z = X$ gives the relationship between entropies roved above, with X and Y interchanged.

Table 1

Propositional notation	Random variable notation	Shannon's notation
H	X	x
E	Y	y
H_i	x	i
E_j	y	j
$P(H_i)$	$P(x)$	P_i
$P(E_j)$	$P(y)$	P_j
$P(H_i \text{ and } E_j)$	$P(x \text{ and } y)$	$p(i, j)$
$P(E_j\|H_i)$	$P(y\|x)$	$p_i(j)$
ent $(\mathbf{H})$	ent X	$H(x)$
ent $(\mathbf{H} \text{ and } \mathbf{E})$	ent $(X \text{ and } Y)$	$H(x, y)$
ent $(\mathbf{E}\|H_i)$	ent $(Y\|x)$	H_i
ent $(\mathbf{E}\|\mathbf{H})$	ent $(Y\|X)$	$H_x(y)$ or H
ent $(\mathbf{H}\|\mathbf{E})$	ent $(X\|Y)$	$H_y(x)$
nt $(\mathbf{H}:\mathbf{E})$ = ent $(\mathbf{E}:\mathbf{H})$ = entropy concerning the H's provided by the method of observation	ent $(X:Y)$ = ent $(Y:X)$ = entropy concerning X provided by the method of observation	R = rate of transmission
ent $(\mathbf{H} : E_j)$	ent $(X : y)$	—
ent $(\mathbf{E} : H_i)$	ent $(Y : x)$	—
ent $(\mathbf{H} : \mathbf{E}\|K)$	ent $(X : Y\|z)$	—
ent $(\mathbf{H} : E_j\|K)$	ent $(X : y\|z)$	—
ent $(\mathbf{H}:\mathbf{E}\|\mathbf{K})$	ent $(X : Y\|Z)$	—

Table 1 shows the equivalence of terms in the propositional tation, the present random variable notation, and Shannon's tation.[20]

(8) CONCLUSION

The notation used in the paper seems to be more self-explanatory and more consistent with the theory of probability (of which information theory is a branch) than other notations in current use. If so, its use would make it easier for mathematicians to communicate with those who wish to apply information theory, and in particular with communication engineers. It is unlikely that this notation will be acceptable to a majority of readers, but its exposition will perhaps bring some extra clarification into information theory.

None of the notation depends on a frequentist theory of probability, though it can all be interpreted in terms of such a theory just as easily as in terms of a subjectivistic one.

Starting with the notion of the amount of information received when we discover that a proposition is true, the development is logically natural and leads to flexible notation and terminology that seems fruitful outside the field of communication theory proper. Support for this view is derived from the various relationships with mathematical statistics that are mentioned. The discussion is of course not intended in any way to detract from Shannon's notable work.

Previous work has been done on the nomenclature of information theory by MacKay.[21] His work may be regarded as complementary to the present paper, since there is hardly any overlapping. Some unpublished works by Lindley[23] on information theory in mathematical statistics, and by McGill and Quastler on terminology have come to light as the paper goes to press; the published works of McGill[24] and McMillan[25] should also be studied.

(9) REFERENCES

(1) JEFFREYS, H.: "Theory of Probability" (Clarendon Press, 1939), p. 29.

(2) POTTER, S.: "Some Notes on Lifemanship" (Rupert Hart-Davis, 1951), p. 30.

(3) VON MISES, R.: "On the Correct Use of Bayes's Formula," *Annals of Mathematical Statistics*, 1942, **13**, p. 156.

(4) GOOD, I. J.: Contribution to the discussion on "The Concept of Probability," *Journal of the Institute of Actuaries*, 1954, **80**, p. 19.

(5) FISHER, R. A.: "Statistical Methods for Research Workers" (Oliver and Boyd, 1938), p. 320.

(6) WIENER, N.: "Cybernetics" (John Wiley, New York, 1948), p. 75.

(7) GOOD, I. J.: "Probability and the Weighing of Evidence" (Charles Griffin, 1950), pp. 2 and 75.

(8) HILBERT, D., and ACKERMANN, W.: "Grundzüge der Theoretischen Logik" (Springer, 1928, English translation: Chelsea, New York, 1950), p. 3.

(9) SAVAGE, L. J.: "The Foundations of Statistics" (Chapman and Hall, 1954).

(10) WOODWARD, P. M.: "Probability and Information Theory, with Applications to Radar" (Pergamon Press, London, 1953).

(11) GOLDMAN, S.: "Information Theory" (Constable, 1953), p. 4.

(12) GOOD, I. J.: "On the Estimation of Small Frequencies in Contingency Tables," *Journal of the Royal Statistical Society, B.*

(13) JAHNKE, E., and EMDE, F.: "Funktionentafeln mit Formeln und Kurven" (Teubner, Leipzig, 2nd Edition, 1933), p. 112.

(14) GOOD, I. J.: "Rational Decisions," *Journal of the Royal Statistical Society, B*, 1952, **14**, p. 107.

(15) GOOD, I. J.: "The Appropriate Mathematical Tools for Describing and Measuring Uncertainty," Chapter 3 of "Uncertainty and Business Decisions," British Associa-

tion Symposium, 1953 (Liverpool University Press, 1954), p. 19.

(16) Jeffreys, H.: "Further Significance Tests," *Proceedings of the Cambridge Philosophical Society*, 1936, **42**, p. 239.

(17) Bartlett, M. S.: "The Statistical Significance of Odd Bits of Information," *Biometrika*, 1952, **39**, p. 230.

(18) Good, I. J.: "On the Population Frequencies of Species and the Estimation of Population Parameters," *ibid.*, 1953, **40**, p. 237.

(19) Good, I. J.: Contribution to discussion on "The Statistical Approach to the Analysis of Time-Series," by M. S. Bartlett (Symposium on Information Theory, Ministry of Supply, 1950), p. 180.

(20) Shannon, C. E.: "The Mathematical Theory of Communication," *Bell System Technical Journal*, 1948, **27**, pp. 379 and 623.

(21) Mackay, D. M.: "The Nomenclature of Information Theory" (Symposium on Information Theory, Ministry of Supply, 1950), p. 13.

(22) Good, I. J.: Contribution to the Discussion on "The Theory of Information," *Journal of the Royal Statistical Society, B*, 1951, **13**, p. 61.

(23) Lindley, D. V.: "On a Measure of the Information provided by an Experiment" (unpublished).

(24) McGill, W. J.: "Multivariate Information Transmission," *Psychometrika*, 1954, **19**, p. 97.

(25) McMillan, B.: "The Basic Theorems of Information Theory," *Annals of Mathematical Statistics*, 1951, **24**, p. 196.

(26) Kullback, S., and Leibler, R. A.: "On Information and Sufficiency," *ibid.* p. 79.

21.372.8: 621.318.013

Monograph No. 156 R
Dec. 1955

ATTENUATION AND PERMEABILITY OF FERROMAGNETIC WAVEGUIDES BETWEEN 9 000 AND 9 675 Mc/s

By J. ALLISON, B.Sc.(Eng.), Ph.D. Graduate, and F. A. BENSON, M.Eng., Ph.D., Associate Member.

The paper was first received 3rd July, and in revised form 14th September, 1955. *It was published as an* INSTITUTION MONOGRAPH *in December,* 1955.)

SUMMARY

Measurements of the attenuations produced by air-filled rectangular waveguides of nickel, mild steel, Mumetal, Radiometal and Rhometal have been made in the frequency range 9 000–9 675 Mc/s. The permeabilities of the materials have been determined from these measurements and a knowledge of the roughness and resistivity of each waveguide internal surface. The effects of temperature on the h.f. permeabilities have also been studied, and some qualitative results are included on the effect of superimposing a steady magnetic field on the h.f. one.

LIST OF PRINCIPAL SYMBOLS

a = Short internal dimension of waveguide.
b = Long internal dimension of waveguide.
c = Velocity of light.
K_{T1}, K_{T2}, Kp = Surface-roughness factors in attenuation formula.
σ = Conductivity of waveguide wall metal.
μ_R = Permeability of waveguide wall metal.
λ_e = Wavelength in unbounded dielectric.
λ_{cr} = Critical guide wavelength.
λ_g = Guide wavelength.
α = Attenuation produced by wall metal.

(1) INTRODUCTION

Many investigations have been made of the dispersion of apparent magnetic permeability with frequency. The results of such measurements are of great value for understanding the elementary magnetization processes by providing more information in connection with the domain-structure theory of ferromagnetic materials.

In 1903 Hagen and Rubens[1] deduced, from measurements of reflection coefficient, that in the frequency range from 10^{13} c/s to 3×10^{14} c/s the permeability of iron is unity, and in 1919 Arkadiew[2, 3] calculated the permeabilities of iron, steel and nickel wires at frequencies of 2 910, 2 970 and 22 900 Mc/s.

In 1945 Allanson[4] comprehensively reviewed and critically discussed the majority of measurements dealing with the permeability of ferromagnetic materials, including alloys and dusts, at frequencies ranging from 100 kc/s to 10 000 Mc/s. At high frequencies the permeability of a specimen may be obtained by methods which fall into two distinct classes. In the first, a measurement is made of the resistive losses in a circuit containing the ferromagnetic material, while the second depends on the effective reactance of a similar circuit. The permeability deduced from measurements in the first group is generally denoted by μ_R and that from the second category by μ_L. It has not always been realized, as Kittel[5] has pointed out, that the two different types of measurement inherently disclose different aspects of the same physical phenomena. The difference between μ_R and μ_L can be treated formally by considering the permeability to be a complex quantity. Thus, the impedance of a circuit element containing a ferromagnetic material may be defined[5] as

$$Z_{cal}(\mu, f) = R_{cal}(\mu, f) + jX_{cal}(\mu, f)$$

If the results of a series of measurements on a particular circuit element give experimental values

$$Z_{exp}(f) = R_{exp}(f) + jX_{exp}(f)$$

then the effective permeability can be defined as the value of μ which makes

$$Z_{cal}(\mu, f) = Z_{exp}(f)$$

Therefore μ will generally be a function of the frequency f and will be complex.

If only $R_{exp}(f)$ is measured, the permeability is taken to make $R_{cal}(\mu_R, f) = R_{exp}(f)$, this relation determining the real function $\mu_R(f)$. Similarly, if only $X_{exp}(f)$ is measured, the real function $\mu_L(f)$ is determined by

$$X_{cal}(\mu_L, f) = X_{exp}(f)$$

There is no simple and direct connection between μ_R and μ_L and the complex permeability, but Millership and Webster[6] have developed expressions connecting these variables for the specific case of a coaxial transmission line which enables the complex permeability to be calculated from measurements of μ_R and μ_L at the same frequency. When μ_R and μ_L are determined for the same specimen it is found that μ_R is larger than μ_L. Dispersion in μ_L occurs at frequencies well below those at which the variation of μ_R is greatest. In fact, preliminary investigations by Millership and Webster[6] indicated that for iron, steel and Mumetal μ_L was unity at 1 500 Mc/s. It is the resistive permeability μ_R which has been measured during the investigations described in this paper.

Since 1945 several investigators have made measurements on ferromagnetic materials in the microwave region.[5–12] Hodsman,[7] Millership,[6, 7] Eichholz[7] and Webster[6] have found the resistive permeabilities of wires at frequencies from 2 290 to 10 084 Mc/s by comparing the attenuations of a coaxial transmission line with first a ferromagnetic specimen and then a non-ferromagnetic reference material as the inner conductor. Millership and Webster[6] have also determined the inductive permeabilities for various materials from a measure of the wavelength in the line. The majority of other recent measurements have been performed at one particular frequency and not over a wide band. Simon[8, 9] has given figures for the permeabilities of nickel films and wires in the range 1 500–9 375 Mc/s, and Eichholz and Hodsman[10] have concluded from theoretical and experimental studies of the reflection of microwaves by a ferromagnetic plate or film that the permeability cannot readily be obtained in this way. Senyal and Chatterjee[11] have deduced the apparent permeabilities of soft iron and nickel plates at 9 375 Mc/s by comparing the resonant fre-

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Allison and Dr. Benson are at the University of Sheffield.

quency and quality factor of a cylindrical cavity with non-magnetic and magnetic end-plates.

It has been shown previously by one of the authors[12] that attenuation measurements on ferromagnetic waveguides can give accurate values for the high-frequency permeabilities of the materials used, provided that the roughnesses of the internal surfaces can be estimated with reasonable certainty. An attempt has also been made[12] to determine the permeabilities of nickel and mild steel at 9 375 Mc/s from attenuation measurements on waveguides of these materials. Prior to this work permeability had not been determined from measurements on a ferromagnetic waveguide, with the exception of some calculations made by Kittel[5] from figures obtained by Maxwell[13] at 24 000 Mc/s on electroplated iron, cold-rolled steel and nickel. It has even been stated that this method is unsuitable for determining permeability at frequencies lower than about 24 000 Mc/s.[7]

The wide variation between the results of the individual investigators and the consequent need for further measurements has prompted the present investigations. Measurements of the attenuations produced in ferromagnetic waveguides have enabled the permeabilities of nickel, mild steel, Mumetal, Rhometal and Radiometal to be calculated over the relatively narrow band of frequencies 9 000–9 675 Mc/s. In these calculations the necessary allowance has been made for internal roughness of the waveguide surfaces. The effect of temperature on the h.f. permeability of these metals has also been studied, and some qualitative results have been found on the effect of superimposing a unidirectional magnetic field on the h.f. one.

(2) CALCULATION OF ATTENUATION AND PERMEABILITY OF FERROMAGNETIC WAVEGUIDES

It has been shown previously by the authors[14] that the attenuation in an air-filled rectangular waveguide carrying the normal H_{01} mode and having irregularities on its surfaces which are, in general, much greater than the skin depth can be expressed as:

$$\alpha = \left(\frac{c}{\sigma}\mu_R\right)^{1/2} \frac{\lambda_g}{b(\lambda_e)^{3/2}} \left[\left(K_{T2} + \frac{b}{2a}K_{T1}\right)\frac{\lambda_e^2}{\lambda_{cr}^2} + Kp\frac{b}{2a}\left(1 - \frac{\lambda_e^2}{\lambda_{cr}^2}\right) \right] \quad . \quad . \quad . \quad . \quad (1)$$

The surface roughness factors K_{T1}, K_{T2} and Kp are the ratios of actual to ideal surface lengths for the long and short sides transverse to the axis and in the longitudinal direction respectively. It is therefore evident that, if the attenuation in a length of ferromagnetic waveguide is measured at a given frequency and the conductivity σ and the roughness coefficients K_{T1}, K_{T2} and Kp are found, the permeability of the wall metal, μ_R, can be calculated provided that the assumption made in deriving eqn. (1) about the magnitudes of the irregularities is valid.

If the depths of the surface irregularities in a specific case are small compared with the skin depth, the surface finish would be expected to have very little effect on the losses. A theoretical study of the power dissipated by eddy currents in a metallic surface at microwave frequencies in the presence of regular parallel grooves or scratches, whose dimensions are comparable to the skin depth, has been made by Morgan.[15] He concludes that the power dissipation in corrugations of various sizes and shapes transverse to the direction of induced-current flow is increased by about 60% over its value for a smooth surface when the r.m.s. deviation of the grooved surface from an average plane is equal to the skin depth. The exact shape of the grooves, according to Morgan, is not critical. Loss caused by grooves, parallel to the current flow is shown, in a particular case, to be about one-third as great as the increase caused by transverse grooves of similar size. The special kind of surface studied by Morgan, however, has never been approximated to by any of the large number of surfaces examined by the authors during the present and previous[14] investigations, and it seems that it would rarely be found in practice.

(3) EXPERIMENTAL PROCEDURE

(3.1) Attenuation Measurements

The attenuations produced in nickel, mild steel, Mumetal, Radiometal and Rhometal waveguides have been measured using the arrangement shown schematically in Fig. 1. The method is similar to that already described by one of the authors,[16] the attenuation per unit length being determined from voltage standing-wave ratio (v.s.w.r.) measurements, although certain modifications have now been introduced. A reflex klystron (type CV129) was used, thus allowing measurements to be made over the frequency range 9000–9 675 Mc/s. The grid of the klystron is amplitude modulated by a tunable square-wave generator operating at about 1 250 c/s. The rectified signal from the standing-wave indicator is amplified by a selective amplifier, whose pass frequency is identical with the frequency of the modulator, and displayed on an indicating meter. This system

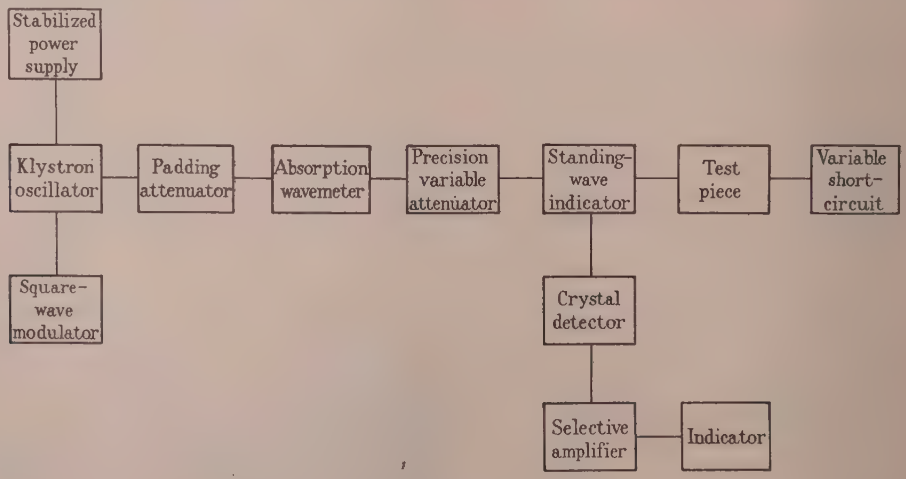


Fig. 1.—Schematic of measuring equipment.

s very sensitive and enables small attenuations to be measured, hus eliminating the need for excessively long lengths of waveuide which were needed for the earlier work.[16]

The v.s.w.r. of the test sample is found as follows. First, the elative amplitudes of the minima are found as the variable shortircuit is moved over a distance λ_g. Then the v.s.w.r. is measured vith the short-circuit in the position giving the minimum of least mplitude and again with the short-circuit piston moved a listance of $\lambda_g/4$ from its first position. The v.s.w.r.'s may be neasured either by the Roberts and von Hippel method[17, 18] or oy a modification of this employing a precision variable attenuator, as described by Vogelman.[19] The attenuation is then the nean of those calculated from the v.s.w.r.'s at the two points. From this is deduced the attenuation produced by the shortircuit alone, which is found by the same method.

The ferromagnetic samples examined were as follows:

(*a*) Commercially-pure-nickel drawn waveguide.
(*b*) Electroplated nickel on precision-drawn brass waveguide.
(*c*) Mild-steel waveguide carefully machined from the solid and joined down the centre of the long sides.
(*d*) Mumetal, Radiometal and Rhometal waveguides fabricated from 0·015 in sheet with junctions along the centres of the long sides, a method which is now being used commercially for producing lightweight waveguides.

The internal dimensions of all the waveguides were 1 in × ½ in.

(3.2) Surface-Roughness Measurements

The surface finish of each waveguide was observed to compare he magnitudes of the irregularities with the corresponding skin lepths. It was confirmed that, in general, the skin depths were mall compared with the dimensions of the irregularities, so xact values of the coefficients K_{T1}, K_{T2} and Kp for all the specimens were found using the microscopic technique developed by he authors.[14] Such measurements are necessary if the precise alues of permeabilities are to be calculated from the attenuation neasurements using eqn. (1). Even in the Mumetal waveguide vith a skin depth of over 100 microinches at a frequency of 675 Mc/s, the irregularities, which were few, were still larger han the skin depth.

(3.3) Resistivity Measurements

Accurate figures for the d.c. resistivities of the waveguide naterials used are also essential in order to calculate the permeabilities from eqn. (1). The resistivity has therefore been determined for each material as accurately as possible by measuring the resistance of a known length of waveguide and ascertaining its cross-sectional area by a method previously described.[16] The resistivity of each material has also been found at temperatures up to 1 000° C, since this information is required for the determination of the variation of permeability with temperature.

(3.4) Temperature Effects

To study the effects of increase of temperature on the attenuations and permeabilities of the various ferromagnetic waveguides, the attenuation of a short length of each guide has been measured as its temperature was increased from room temperature to 1 000° C in a muffle furnace. Cooling of the remainder of the measuring equipment was effected by the use of subsidiary nonmagnetic waveguide connecting junctions whose attenuation was subsequently subtracted from the total attenuation, giving that of the magnetic sample alone.

(3.5) Effect of External Superimposed D.C. Field

The effect of superimposing a unidirectional magnetizing field on the h.f. field has been studied by making the test piece one arm of a rectangular magnetic yoke. The current through a coil of many turns placed on one of the other limbs of the yoke may be changed so as to provide a varying flux-density in the sample. After each specimen had been demagnetized by passing a decaying alternating current through the magnetizing coil, its attenuation was measured as the direct current through the coil was varied.

(4) RESULTS

Figures for the attenuation produced by the various specimens at 9 300 Mc/s, together with the measured d.c. resistivities and surface-roughness factors, are given in Table 1. Allanson[4] has pointed out that it is difficult to compare the results of many of the early investigators because of wide variations in composition, heat treatment and mechanical handling of the materials used. Thus in Table 1, details of the analysis and initial permeability of each sample, except electrolytic nickel, are included.

The changes in permeability with frequency, temperature and polarizing d.c. field have been calculated for each specimen from

Table 1

Characteristics of the Various Ferromagnetic Materials Examined

Material	Analysis	D.C. resistivity	Initial permeability	Attenuation at 9 300 Mc/s	K_{T1}	K_{T2}	Kp	Remarks
		microhm-cm		dB/m				
Nickel ..	Contains following impurities: Fe = 0·58% S = 0·022% Mn = 0·12% Si = trace	9·39	13·5	0·394	1·119	1·109	1·090	Drawn tube
Electrolytic nickel ..	—	10·3	—	0·337	1·010	1·007	1·007	Electrodeposited on precision-drawn brass waveguide
Mild-steel ..	Contains following impurities: C = 0·27% Si = 0·03% S = 0·053% P = 0·036% Mn = 0·74% Ni = 0·15%	13·5	141	1·435	1·026	1·013	1·009	Milled from solid bar
Mumetal ..	Ni = 72·2%* Cu = 5·3%	65·3	21 000†	0·841	1·002	1·002	1·001	Fabricated from 0·015 in sheets; then annealed *in vacuo* at 1050° C for 2 hours and cooled to 200° C in 8 hours
Radiometal	Ni = 49·1%* Cu = 0·0	46·6	1 850†	1·345	1·002	1·002	1·001	
Rhometal ..	Ni = 37·5%* Cu = 0·0	74·2	1 200†	1·66	1·002	1·002	1·001	

* Remainder iron with traces of molybdenum and chromium.
† Manufacturer's figures.

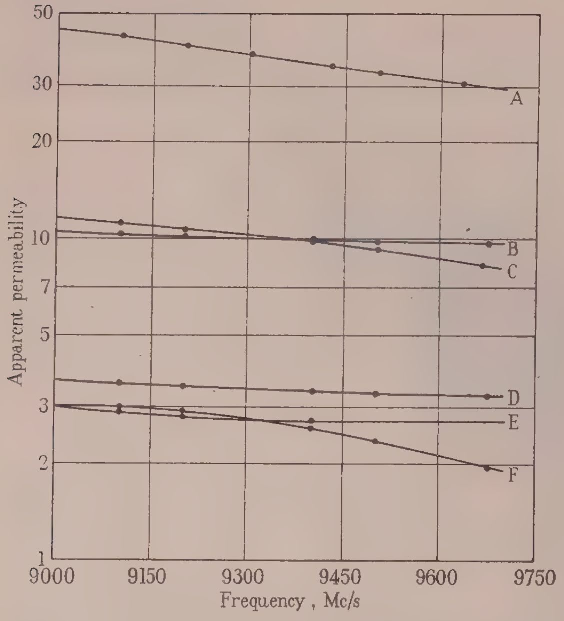


Fig. 2.—Variation of apparent permeability with frequency.

A = Mild steel.
B = Rhometal.
C = Radiometal.
D = Nickel.
E = Electrolytic nickel.
F = Mumetal.

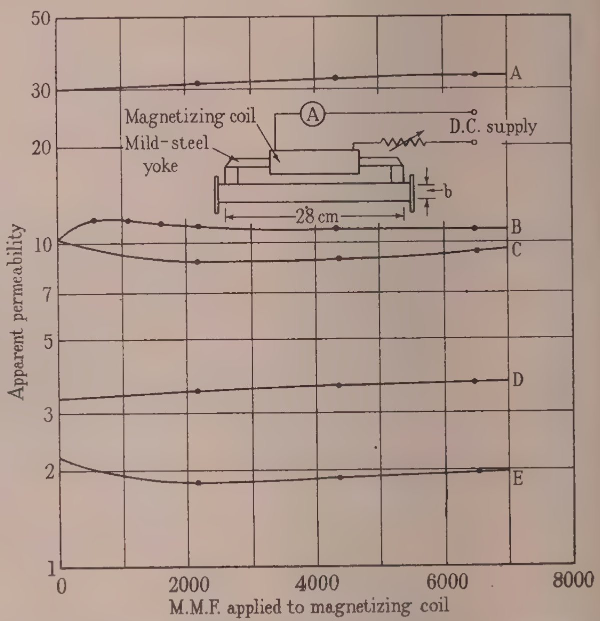


Fig. 4.—Effect of superimposed d.c. field on apparent permeability.

A = Mild steel.
B = Rhometal.
C = Radiometal.
D = Nickel.
E = Mumetal.

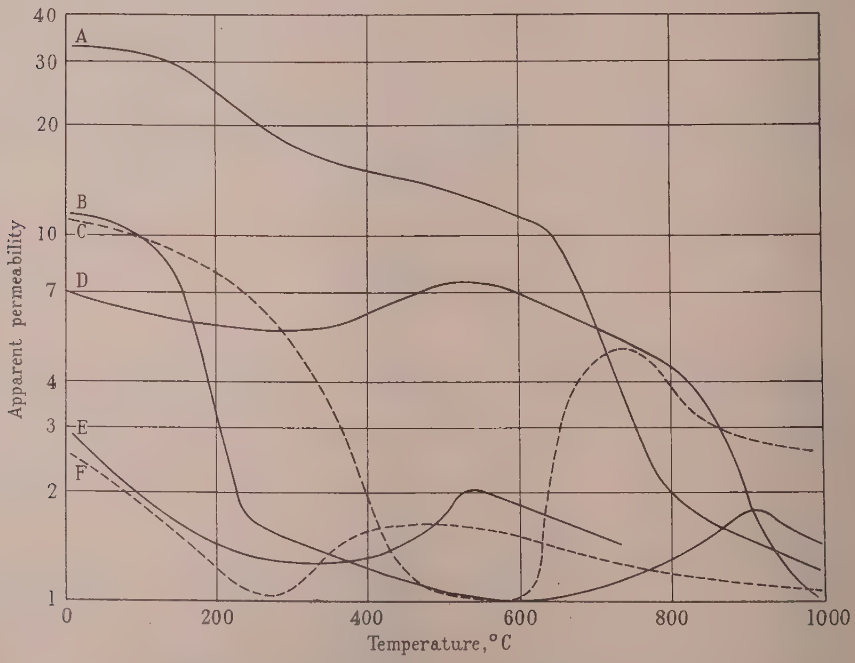


Fig. 3.—Variation of permeability with temperature.

A = Mild steel.
B = Rhometal.
C = Radiometal.
D = Nickel.
E = Electrolytic nickel.
F = Mumetal.

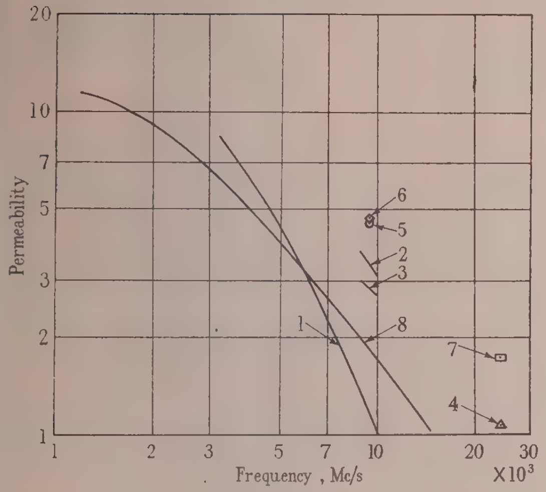


Fig. 5.—Variation of permeability μ_R of nickel with frequency.

1) Hodsman, Eichholz and Millership:[7] nickel wire.
2) Authors' results: drawn waveguide.
3) Authors' results: electroplated waveguide.
4) Maxwell[13] (as publisbed by Kittel[5]): electroformed waveguide.
(5) Simon:[9] nickel wire.
(6) Senyal and Chatterjee:[11] thick plates.
(7) Maxwell[13] (as published by Kittel[5]): electroplated waveguide.
(8) Arkadiew:[2] nickel wire.

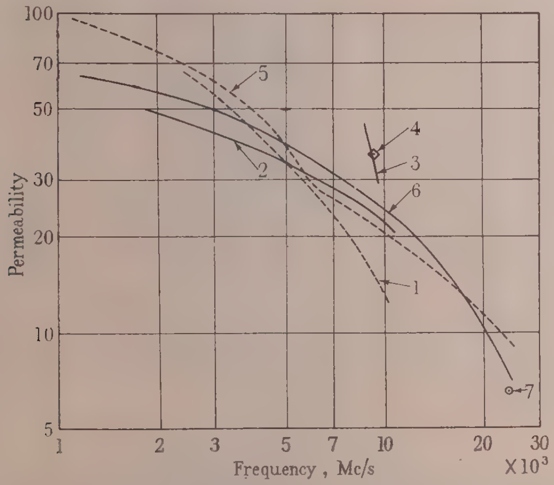


Fig. 6.—Variation of permeability μ_R of iron and steel with frequency.

1) Hodsman, Eichholz and Millership:[7] bright-steel wires.
2) Hodsman, Eichholz and Millership: annealed-steel wires.
3) Authors' results: mild-steel waveguide.
(4) Senyal and Chatterjee:[11] soft-iron thick plates.
(5) Arkadiew:[2] mild-steel wires.
(6) Arkadiew:[2] Swedish-iron wires.
(7) Maxwell:[13] (as published by Kittel[5]): cold-rolled-steel waveguide.

he attenuation, d.c. resistivity and surface-roughness measurements, using eqn. (1). These results are shown graphically in Figs. 2, 3 and 4.

Figs. 5 and 6 give the majority of the previously published esults for the resistive permeabilities of nickel and iron and steel at frequencies around 10 000 Mc/s, together with the present esults for comparison.

(5) DISCUSSION OF RESULTS

(5.1) Permeability-Frequency Curves

It is now well known that the permeability of a ferromagnetic naterial falls considerably in the microwave region. All the metals examined show some degree of dispersion of permeability over the narrow frequency-band employed. It appears from the results that for the materials tested the higher the initial permeability the greater is the relative decrease of μ_R from its d.c. value to its value in the 3 cm band. There seems to be no simple connection, however, between the initial permeability and the dispersion over the range 9 000 to 9 675 Mc/s. The results now obtained are rather higher than the average, a fact which was also mentioned in earlier work on nickel and mild-steel waveguides.[12] There is probably some significance in the fact that, for both nickel (with the exception of Simon's result) and steel, the lower values have been obtained from experiments on wires whilst the comparable higher values have been acquired from measurements on thick plates. This is thought to be largely due to the different treatments to which the machined plate and drawn-waveguide surfaces have been subjected compared with wires. Incidentally, Simon's results for nickel wires are probably not very reliable, because the samples were used as bolometers and were therefore heated during test.

Becker[20] attributes the dispersion mentioned above to microscopic eddy currents which retard the rapidly moving domain walls. Eventually a frequency is reached when the induced counter field is so large that movement of the walls is prevented altogether and the permeability falls to unity. Millership *et al.*[6,7] have compared their results with the Becker theory and find that, although the general forms of the experimental and theoretical curves are similar, the real part of the permeability decreases more rapidly than the theory indicates. Smidt[21] has carried out similar experiments using a coaxial line with a magnetic inner conductor and he, too, finds that the dispersion of μ_L is greater than the theory predicts.

A theory has been developed by Kittel[5] which ignores eddy-current effects and attributes the dispersion solely to the depth of penetration of the field at very high frequencies being comparable with the thickness of the ferromagnetic domains, which has been estimated to be between 10^{-3} and 10^{-4} cm. The effective field acting on the domain boundary is then the mean of the applied field taken over the entire wall, causing a reduction in the apparent permeability. Here again the decrease of permeability with frequency found by experiment is slower than the theory predicts.[5,21]

Döring[22] has used a different approach to the problem and has shown that the domain wall behaves as if it had mass inertia. This inertia effect is larger than the eddy-current effect suggested by Becker and may play a more important part in the dispersion. Recently, Kittel[23] and Birks[24] have shown that the damping term associated with spin precession in ferromagnetic resonance experiments is even more important than the inertia effects calculated by Döring and now appears to be the principal reason for the decrease of initial permeability with frequency. The Becker and Kittel theories suffer from the facts that, on one, a simple cubic domain model has been chosen on which to base the calculations, whereas in the other a film one domain in thickness is considered. A further objection to the Becker theory in the microwave règion is that no account is taken of the reduced depth of field. Further, Kittel assumes that the domain boundary moves as a rigid whole, when there is no reason why it should not yield locally, so providing bending instead of movement.

It should also be remembered that the domain structure of the surface layer is different from that of the body of the material giving varying magnetic properties with skin depth, and therefore frequency, if boundary movements are responsible for magnetization at high frequencies. That the surface consists of triangular domains preventing the appearance of free poles at the surface has, in fact, been suggested by Néel,[25] whereas in the body of the material the domains will probably be in the form of plane parallel sheets. It has already been mentioned that the permeability at high frequencies is likely to depend very much on the physical character of the surface. At 10^5 c/s, for example, thin films of oxide formed during heat treatment have been found[26] to cause the apparent permeability to decrease by a factor of 10. It has even been suggested that dispersion is due to the presence of surface layers of non-ferromagnetic material,[27] but the existence of such layers is now thought to be unlikely.[6]

The strength of the magnetic field applied to the specimen with the type of measurement described here is very small (of the order of 0·01 oersted). Thus the permeabilities calculated refer to magnetization at the commencement of the magnetization cycle.

(5.2) Permeability/Temperature Curves

Bozorth[28] records that if the field strength is low the initial d.c. permeability of a material increases with temperature but finally falls to unity at the Curie point. Glathart[29] obtained similar results at a frequency of 200 Mc/s. Although the authors could not find any reports of investigations at higher frequencies it seemed likely at the commencement of the present work that similar results could be expected in the microwave region.

Because of the impossibility of measuring accurately the loss in very short lengths of waveguide, the actual length of each sample examined was comparable with that of the heating furnace. Thus there is a possibility of a variation of temperature along the length of a specimen and there may be some error in the actual temperatures recorded. Consequently, the sharp changes occurring in μ_R may be even more severe. It is therefore evident that the temperature curves could be displaced slightly, indicating, for example, that the permeability of the mild-steel sample changes suddenly in the neighbourhood of the Curie point (770° C), falling to unity very rapidly. There is, however, no initial increase of permeability with increasing temperature as at lower frequencies.

It seems evident that, at temperatures up to about 400–500° C, the changes in permeability are due to changes in the metallic state of the material. It has been suggested that the initial changes in permeability may be due to the relieving of stresses in the material, since all the metals, except the drawn-nickel specimen, were taken through only one heat run. This tends to be disproved by the fact that although the drawn-nickel waveguide was taken through two heat cycles before measurements were made, with a consequent considerable increase of its initial apparent permeability at room temperature, the general shape of the μ_R/temperature curve is still the same as that for the electrolytic nickel. There is also a possibility that the initial decrease of permeability may be due to absorption of gas from the atmosphere or, in electrolytic nickel, the liberation of hydrogen. At temperatures over about 400–500° C oxide films formed on the metal surfaces will affect the values obtained for μ_R. These oxides (iron oxide being the dominant one), apart from causing the surface to roughen, have a resistivity which decreases as the temperature is increased.

(5.3) Effect of Superimposed D.C. Field

Harrison *et al.*[30] have applied a d.c. axial field to a Mumetal wire carrying a 55 c/s alternating current and have found that there is a decrease in maximum permeability, the change being 60% for an applied field of 0·6 oersted. Whilst changes of such magnitude do not appear to occur at higher frequencies, the external field still influences the properties of the material to some extent as may be seen from Fig. 4. Owing to leakage effects it is very difficult to estimate the actual magnetic field strength in the specimen. Consequently the circuit employed is shown in Fig. 4 and apparent permeability is plotted against magnetizing-coil m.m.f. Over the range used there are deviations of up to 20% in the permeability. Presumably the initial shapes of the magnetization curves for the materials are largely responsible for the changes in permeability recorded.

(6) GENERAL CONCLUSIONS

The dispersion of permeability in the microwave band is of considerable value for understanding the elementary magnetization processes, and although it has been studied extensively the results of the individual investigations have shown large variations. It does seem, however, from a comparison of the present and previous studies that—because the skin-depth at these high frequencies is small and therefore the permeability is a property of the surface layer of a material—the discrepancies are probably largely due to the different treatments to which the various surfaces of the samples have been subjected. It also now looks as though the dispersion effects at microwave frequencies are associated with spin precession rather than with eddy currents, reduced effective field strength or mass inertia of domain walls, as previously suggested by several workers. It is interesting to note that the permeability of most of the materials examined will not apparently reach unity until the frequency is much greater than 9 700 Mc/s.

The present work has demonstrated that the losses in a ferromagnetic waveguide are quite high because of the values of permeability and effective resistivity, and it has been confirmed that attenuation measurements on such a waveguide will give an accurate value for the high-frequency permeability of the material used provided that precise information is available about the roughness of the internal surfaces. It is difficult to establish the absolute degree of accuracy of the method. The source of largest error is the attenuation measurements, but the probably error in each attenuation value is estimated to be less than 2%. The surface-roughness measurements need to be limited to a small number of cross-sections, but experience has shown repeatedly that different sections from the same length of waveguide give surprisingly similar results.

It is evident that there is no simple explanation for the complex nature of the permeability/temperature curves, but the variations observed are apparently connected with changes in lattice structure, relief from stress and oxidation of the surface of the metal as the temperature increases.

(7) ACKNOWLEDGMENTS

The work recorded in the paper has been carried out in the Department of Electrical Engineering at the University of Sheffield. The authors wish to thank Mr. O. I. Butler for facilities afforded in the laboratories of this Department. They are also grateful to The Telegraph Construction and Maintenance Co. Ltd., for supplying samples of Mumetal, Radiometal and Rhometal sheets.

(8) REFERENCES

(1) Hagen, E., and Rubens, H.: "Über Beziehungen des Reflexions- und Emissionsvermögens der Metalle zum elektrischen Leitvermögen," *Annalen der Physik*, 1903, **11,** p. 873.

(2) Arkadiew, W.: "Über die Absorption elektromagnetischer Wellen an zwei parallelen Drähten," *ibid.*, 1919, **58,** p. 105.

(3) Arkadiew, W.: "Über die Reflexion elektromagnetischer Wellen an Drähten," *ibid.*, 1914, **45,** p. 133.

(4) Allanson, J. T.: "The Permeability of Ferromagnetic Materials at Frequencies between 10^5 and 10^{10} c/s," *Journal I.E.E.*, 1945, **92,** Part III, p. 247.

(5) Kittel, C.: "Theory of the Dispersion of Magnetic Permeability in Ferromagnetic Materials at Microwave Frequencies," *Physical Review*, 1946, **70**, p. 281.

(6) Millership, R., and Webster, F. V.: "High-Frequency Permeability of Ferromagnetic Materials," *Proceedings of the Physical Society*, B, 1950, **63**, p. 783.

(7) Hodsman, G. F., Eichholz, G., and Millership, R.: "Magnetic Dispersion at Microwave Frequencies," *ibid.*, 1949, **62**, p. 377.

(8) Simon, I.: "Magnetic Permeability of Nickel in the Region of Centimetre Waves," *Nature*, 1946, **157**, p. 735.

(9) Simon, I.: "Measurement of Permeability and Ferromagnetic Resonance at Microwave Frequencies," *Časopis pro Pěstování Mathematiky a Fysiky*, 1948, **73**, p. 41.

10) Eichholz, G., and Hodsman, G. F.: "Reflection of Microwaves by Ferromagnetic Materials," *Proceedings of the Leeds Philosophical and Literary Society*, 1946, **4**, p. 303.

11) Senyal, G. S., and Chatterjee, J. S.: "Measurement of Ferromagnetic Permeability at Microwave Frequencies," *Indian Journal of Physics*, 1953, **27**, p. 238.

12) Benson, F. A.: "Attenuation in Nickel and Mild-Steel Waveguides at 9 375 Mc/s," *Proceedings I.E.E.*, Paper No. 1577 R, January, 1954 (**101**, Part III, p. 38).

13) Maxwell, E.: "Conductivity of Metallic Surfaces at Microwave Frequencies," *Journal of Applied Physics*, 1947, **18**, p. 629.

14) Allison, J., and Benson, F. A.: "Surface Roughness and Attenuation of Precision-Drawn, Chemically Polished, Electropolished, Electroplated and Electroformed Waveguides," *Proceedings I.E.E.*, Paper No. 1785 R, March, 1955 (**102 B**, p. 251).

15) Morgan, S. P.: "Effect of Surface Roughness on Eddy Current Losses at Microwave Frequencies," *Journal of Applied Physics*, 1949, **20**, p. 352.

16) Benson, F. A.: "Waveguide Attenuation and its Correlation with Surface Roughness," *Proceedings I.E.E.*, Paper No. 1467 R, March, 1953 (**100**, Part III, p. 85).

(17) Roberts, S., and von Hippell, A.: "A New Method of Measuring Dielectric Constant and Loss in the Range of Centimetric Waves," *Journal of Applied Physics*, 1946, **17**, p. 610.

(18) Barlow, H. M., and Cullen, A. L.: "Microwave Measurements" (Constable, London, 1950), Chapter 5.

(19) Vogelman, J. H.: "Precision Measurements of Waveguide Attenuation," *Electronics*, 1953, **26**, p. 196.

(20) Becker, R.: "Ferromagnetismus bei Hochfrequenten Wechselfeldern," *Zeitschrift für Technische Physik*, 1938, **19**, p. 542.

(21) Smidt, J. P.: "High Frequency Permeability," *Applied Scientific Research*, 1950, **B1**, p. 127.

(22) Döring, W.: "Über die Trägheit der Wände zwischen Weiss'schen Bezirken," *Zeitschrift für Naturforschung*, 1948, **3a**, p. 373.

(23) Kittel, C.: "Ferromagnetic Resonance," *Journal de Physique et le Radium*, 1951, **12**, p. 291.

(24) Birks, J. B.: "Properties of Ferromagnetic Compounds at Centimetre Wavelengths," *Proceedings of the Physical Society*, B, 1950, **63**, p. 65.

(25) Néel, L.: "Laws of Magnetism and Subdivision of Elementary Domains of Iron," *Journal de Physique et le Radium*, 1944, **5**, p. 241.

(26) Peterson, E., and Wrathall, L. R.: "Eddy Currents in Composite Laminations," *Proceedings of the Institute of Radio Engineers*, 1936, **24**, p. 275.

(27) Wein, M.: "Über die Hautwirkung ferromagnetischer Drähte bei Hochfrequenz," *Annalen der Physik*, 1931, **8**, p. 899.

(28) Bozorth, R. M.: "Ferromagnetism" (D. Van Nostrand, New York, 1951), p. 714.

(29) Glathart, J. L.: "The Inner, Initial Magnetic Permeability of Iron and Nickel," *Physical Review*, 1939, **55**, p. 833.

(30) Harrison, E. P., Tuney, G. L., Rowe, H., and Gollop, P. H.: "The Electrical Properties of High Permeability Wires carrying Alternating Current," *Proceedings of the Royal Society*, A, 1936, **157**, p. 451.

621.3.013 : 621.313.3

Monograph No. 157 U
Dec. 1955

THE THEORY OF THIRD-HARMONIC AND ZERO-SEQUENCE FIELDS

By Professor G. H. RAWCLIFFE, M.A., D.Sc., Member, and B. C. McDERMOTT, B.Sc.Tech., Graduate.

(*The paper was first received* 19*th July, and in revised form* 30*th September*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in December*, 1955.)

SUMMARY

The paper first digests existing information about 3rd-harmonic fields, mainly published previously in piecemeal fashion, and then adapts and extends it into modern form for use in symmetrical-component theory as applied to machines. This theory has provided the basis for a very successful 3 : 1 pole-changing machine which is described in another paper,[8] and the present paper includes a number of test results which verify that the theory of 3rd-harmonic fields given here is essentially sound.

LIST OF SYMBOLS

T_m = Peak number of magnetizing turns acting on pole centre.
I_m = Peak magnetizing current, amp.
h = Peak magnetomotive force for current-vector position I, AT/pole.
n = Number of conductors per pole per phase.
I_a = Current (r.m.s.) per phase, amp.
I = Peak current per phase, amp.
θ = Electrical angle, rad.
m = Order of harmonic.
p = Number of pairs of poles.
q = Number of slots per pole per phase in a.c. winding (primary).
s = Number of slots per pole per phase in d.c. winding (secondary).
V_1 = Generated voltage (r.m.s.) per phase at fundamental frequency, volts.
V_3 = Generated voltage (r.m.s.) per phase at 3rd-harmonic frequency, volts.
α = Angle of short chording (for full pitch $\alpha = 0$).
x_1 = Reactance per phase at fundamental frequency, ohms.
x_3 = Reactance per phase at 3rd-harmonic frequency, ohms.
ϕ_1 = Flux per pole for fundamental pole number, webers.
ϕ_3 = Flux per pole for 3rd-harmonic pole number, webers.
I_1 = Magnetizing current (r.m.s.) for fundamental pole number, amp.
I_3 = Magnetizing current (r.m.s.) for triple pole number, amp.
F_1 = M.M.F. acting on centre of fundamental poles, AT.
F_3 = M.M.F. acting on centre of 3rd-harmonic poles, AT.

Triplen harmonics are harmonics of order $3g$, where $g = 1, 3, 5$, etc.

(1) INTRODUCTION

The 3rd-harmonic space components of a polyphase armature can, in some circumstances, produce a pole-tripling effect, and the magnetomotive force due to any a.c. winding, at one particular instant, can always be simulated by passing direct current through the winding. These two principles can be combined to produce a triple-frequency generator from standard equipment without difficulty. The theory of 3rd-harmonic fields and windings producing them, together with the theoretical discussion of an actual generator including certain test results, forms the first part of the paper.

It is not thought that such a generator is likely to find any serious industrial application, but the collection and consolidation of the theory of 3rd-harmonic fields is a fitting introduction to a description of a new 3 : 1 pole-changing winding of obvious industrial possibilities. This winding, and the performance of a motor incorporating it, are described in a companion paper.[8]

The theory of 3rd-harmonic fields has also recently acquired a new analytical importance. Brown and Butler[4] have described a system of applying the method of symmetrical components to a machine with asymmetrical connections, which could equally be applied to a symmetrical machine with an unsymmetrical supply voltage. This method necessarily involves a knowledge of the positive-, negative- and zero-sequence magnetizing and leakage impedances of the machine under consideration. The zero-sequence impedance of a machine is its impedance when all phases are carrying equal cophasal currents: in effect it is the impedance when series-connected, as in the prototype pole-tripling circuits. Put shortly, the zero-sequence impedance of a machine is the same thing as its basic 3rd-harmonic or triple-pole impedance. If, in any way, the 3rd-harmonic impedance is made equal to zero, as it is if the winding distribution is either sinusoidal or uniformly spread over 120°, the passage of zero-sequence currents will produce no 3rd-harmonic field components. This fact gives new importance to the theoretical analysis of 3rd-harmonic fields, which are therefore considered in detail for a variety of windings.

(2) ADDITION OF POLYPHASE WINDINGS WITH DIRECT-CURRENT EXCITATION

In previous papers, such as those by Parker Smith and Boulding,[1] and by Hague,[2] the m.m.f. due to a polyphase winding has been examined, and a paper by one of the present authors[3] discussed in particular the effect of feeding direct current into a polyphase winding. The methods employed in these papers may be further applied as follows.

Let the fundamental and harmonic space-components of a polyphase winding be represented by

$$T_1 = \Sigma T_m \sin mp\theta$$

$$T_2 = \Sigma T_m \sin m\left(p\theta - \frac{2\pi}{3}\right)$$

$$T_3 = \Sigma T_m \sin m\left(p\theta - \frac{4\pi}{3}\right), \quad (m = 1, 3, 5, 7, \ldots)$$

Suppose that a direct current is fed through the three windings in the manner common in synchronous induction motors,[3] giving effective phase currents $+I_m$, $-I_m/2$ and $-I_m/2$. Then, by simple algebra, the resultant space distribution of m.m.f. can be shown to be

$$\Sigma I_m T_m \sin mp\theta\left(1 + \cos\frac{m\pi}{3}\right) \quad . \quad . \quad . \quad . \quad (1)$$

Correspondence on Monographs is invited for consideration with a view to publication.

Prof. Rawcliffe is Professor of Electrical Engineering, University of Bristol. Mr. McDermott is in the Electrical Engineering Department of the University.

If $m = 1, 5, 7, 11, 13$, etc., the corresponding term is $\frac{3}{2}I_mT_m \sin mp\theta$, but if $m = 3, 9, 15$, etc., it is zero.

Suppose, alternatively, that the three phases are connected in series, in the same sense, and fed with a direct current I_m. Then the resultant space-distribution of m.m.f. can readily be shown to be

$$\Sigma I_mT_m \sin mp\theta\left(1 - 2\cos\frac{m\pi}{3}\right) \quad . \quad . \quad . \quad (2)$$

If $m = 1, 5, 7, 11, 13$, etc., the corresponding term is zero; but if $m = 3, 9, 15$, etc., it is $3I_mT_m \sin mp\theta$.

As is well known, the ordinary method of d.c. excitation, as in a synchronous induction motor, gives all odd harmonics except the third and odd multiples of the third, whereas the series method of d.c. excitation gives only those harmonics which the other method of excitation excludes. Further, in the series method the magnitude of the resultant harmonics is three times the magnitude of their phase components, whereas the magnitude of the resultant harmonics in the normal method of excitation is $1\frac{1}{2}$ times that of the corresponding phase components. These two general expressions, (1) and (2) thus enable the magnitude of the resultant harmonics of m.m.f. due to three windings combined and fed with direct current, in either way, to be calculated from the component harmonic series for each phase taken alone. These principles will be repeatedly applied in the paper.

In a general paper on the m.m.f. of windings, Clayton[5] gave a number of general equations which can be made to cover this case of pole-tripling, but in Section IX of his paper he does not refer to the 3 : 1 pole ratio at all. His treatment throughout is solely mathematical and almost confined to an examination of the m.m.f.'s ideally produced by various windings: the operation of machinery in which such m.m.f.'s may be produced, with regard to load rating, torque characteristics and other aspects, is not discussed.

(3) THIRD HARMONICS AND THE ZERO-SEQUENCE FIELD

A paper by Brown and Butler[4] established, theoretically and practically, the existence of crawling 3rd-harmonic fields in asymmetrically connected induction motors. The method of symmetrical components was used to analyse these fields, and —as these authors put it—the 3rd-harmonic fields are due to the zero-sequence components of the unbalanced currents.

To feed all the phase windings in series, in the same sense, with direct current is in effect to "freeze" the zero-sequence components in a particular vector position, just as the normal methods of feeding direct current into the secondary circuit of a synchronous induction motor are, in effect, methods of freezing the positive-sequence vectors in one of several possible vector positions.[3] The fact that a 3rd-harmonic field can be produced by d.c. series-excitation of a polyphase winding is in fact an elegant extension and practical demonstration of the possibility of producing 3rd-harmonic fields (due to zero-sequence currents) in unbalanced polyphase windings. One is accustomed to suppose that triplen harmonic fields are always absent in 3-phase windings, and therefore it is well to be forcefully reminded that this is only true for a balanced winding carrying balanced currents.

In addition to being of triple space-frequency, the waveform and amplitude of the resultant m.m.f. produced by a 3-phase winding carrying three equal and cophasal currents are very different from the waveform and amplitude of the m.m.f. produced by normal 3-phase currents of the same magnitude flowing in the same winding. Typical m.m.f.'s of triple space-frequency obtained by series excitation are shown in Fig. 3.

The amplitude of the m.m.f. for a given current determines the peak flux per pole, and this flux is governed by the terminal voltage and can be calculated in the usual way. We are thus able successively to calculate the apparent impedance of the winding (per phase) to a 3-phase applied voltage, and its apparent impedance (series-connected) to a single-phase applied voltage. The first impedance is known as the positive-sequence magnetizing impedance per phase and the second as the zero-sequence magnetizing impedance per phase. By symmetry, the negative-sequence magnetizing impedance of a balanced winding will be equal to the positive-sequence magnetizing impedance.

These are impedances of the type discussed by Brown and Butler,[4] and their importance lies in the fact that to examine the behaviour of an unbalanced system it is necessary to resolve the voltages and currents into their balanced symmetrical components, but the impedances presented to the different components are not the same.

In Section 8 detailed calculations are made of the positive- and zero-sequence magnetizing impedances of various balanced 3-phase windings, and these values are compared with the experimental values: in general there is reasonable agreement.

(4) IDEAL VALUE OF TRIPLE-FREQUENCY VOLTAGE

The principle of the triple-frequency generator can be well illustrated, and tested in primitive form, by calculating the triple-frequency voltage generated at a given speed, and comparing it with the fundamental-frequency voltage generated when the generator secondary circuit is excited by the same total direct current, fed into it as though the generator were a synchronous induction motor (s.i.m.).[3] (The usual way of exciting an s.i.m. is by applying d.c. excitation through one phase winding in reverse series with the other two phase windings in parallel.) If V_3 is the triple-frequency voltage for a series-exciting current I_m, and V_1 is the fundamental-frequency voltage for the same exciting current applied as in an s.i.m., the ratio V_3/V_1 for a given speed can be calculated, ignoring saturation, as follows. If the star point only of the secondary circuit is available, the three phase windings can be fed in parallel, the total current being $3I_m$, and the exciting voltage correspondingly reduced. The turns distribution of a single phase of a full-pitch, 60°-spread winding is shown in Fig. 1, and its Fourier analysis can readily be shown to be

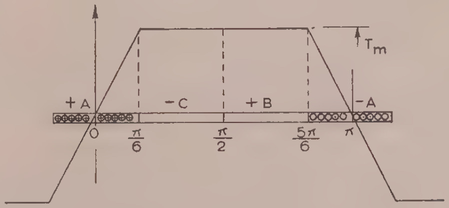


Fig. 1.—Turns distribution of one phase of a 60° spread full-pitch winding.

$$\frac{12}{\pi^2}T_m\left(\sin\theta + \frac{2}{9}\sin 3\theta + \frac{1}{25}\sin 5\theta - \frac{1}{49}\sin 7\theta - \frac{2}{81}\sin 9\theta - \frac{1}{121}\sin 11\theta + \ldots\right) \quad . \quad (3)$$

The 3rd-harmonic m.m.f. content of one phase is therefore two-ninths of the fundamental. For excitation as in a synchronous induction motor, from the above Fourier analysis and the arguments in Section 2, the fundamental flux per pole is therefore

$$K\frac{3}{2}I_mT_m\frac{12}{\pi^2}$$

where K is a constant. This result can otherwise be obtained by taking a Fourier analysis of the resultant m.m.f. of the three phases, which has been shown[1,3] to be

$$2I_mT_m\frac{9}{\pi^2}\left(\sin\theta+\frac{1}{5^2}\sin 5\theta-\frac{1}{7^2}\sin 7\theta-\frac{1}{11^2}\sin 11\theta+\ldots\right) \quad . \quad . \quad . \quad . \quad (4)$$

This expression can also be written down at sight from expression (3) using the methods of Section 2.

For series d.c. excitation, the fundamental flux, which is a 3rd-harmonic flux relative to the flux in the first case, is seen by the arguments of Section 2 to be $3KI_mT_m\frac{12}{\pi^2}\times\frac{2}{9}$.

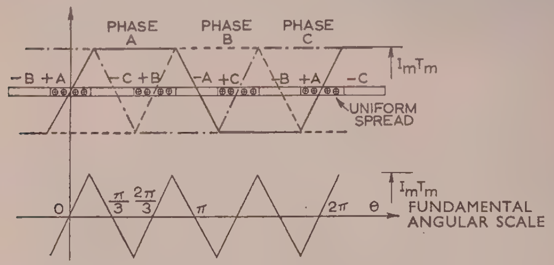


Fig. 2.—M.M.F. distribution of three phases of a 60°-spread full-pitch winding, series-connected and carrying the same current I_m.

Alternatively, the resultant m.m.f. of the three phases excited in series can be seen from Fig. 2 to be of triangular waveform, and from this the Fourier analysis of the resultant m.m.f. is readily obtained as

$$\frac{8}{\pi^2}I_mT_m\left(\sin 3\theta-\frac{1}{3^2}\sin 9\theta+\frac{1}{5^2}\sin 15\theta-\ldots\right) \quad . \quad (5)$$

giving the same result as before for the new fundamental flux. This last Fourier analysis can also be written down at once from expression (3), using the results of Section 2.

The voltages, V_1 and V_3, induced in the primary circuit in each case are proportional to the corresponding fluxes per pole, the conductors per phase, and the spread and chord factors of the primary winding. The spread factor of the primary winding relative to the fundamental flux is, as usual, $3/\pi$; but its spread factor relative to the 3rd-harmonic flux is $\sin\left(\frac{\pi}{2}\right)\Big/\left(\frac{\pi}{2}\right)=\frac{2}{\pi}$, if it is assumed to be uniformly spread. It is assumed, as was true for the machine tested, that the fundamental chord factor, and therefore also the 3rd-harmonic chord factor, was unity.

This gives the theoretical ratio between the 3rd-harmonic and the fundamental voltages for the two methods of d.c. excitation of the secondary circuit, as

$$\left(\frac{\text{3rd-harmonic voltage per phase, } V_3}{\text{Fundamental voltage per phase, } V_1}\right)=\left(\frac{3\times\frac{2}{9}\times\frac{2}{\pi}}{\frac{3}{2}\times 1\times\frac{3}{\pi}}\right)$$

$$=\frac{8}{27}=0\cdot296 \quad . \quad (6)$$

As will be seen in Sections 4 and 5, an approximation to this ratio was obtained in practice, but it was later found necessary to make some allowance for the actual slotting rather than to assume uniform phase spread for the 3rd-harmonic flux and induced voltages.

The fundamental 3rd-harmonic voltages, V_3, in all the phases of the primary will be identical and cophasal, since their angular spacing, $2\pi/3$, with respect to the main flux is equivalent to one pole pair of the 3rd-harmonic flux. Hence the three primary phases can be connected in series when the secondary is series-excited, and we then have

$$\frac{\text{Total 3rd-harmonic voltage}}{\text{Fundamental voltage per phase}}=\frac{8}{9}=0\cdot888 \quad . \quad (7)$$

This also was found to be roughly true in practice, and, moreover, the three voltages were exactly equal and cophasal.

Alternatively, the primary phase windings can be connected in parallel, giving an unaltered primary voltage but three times the current-carrying capacity. It is, of course, a matter of indifference whether stator or rotor is treated as the primary, and for the machine tested, each was so treated in turn. It ought to be emphasized that throughout the paper, in dealing with d.c.-excited machines, the secondary means the winding which is fed with direct current, and the primary the element in which alternating current is generated or to which it is applied.

The machine tested had 3 slots/pole/phase in the stator and 2 slots/pole/phase in the rotor, and normally operated as a 6-pole 400-volt 3-phase 50 c/s induction motor. Both windings were single-layer full-pitch 60° spread concentric windings. The stator phases were completely separate. The rotor phases were con-

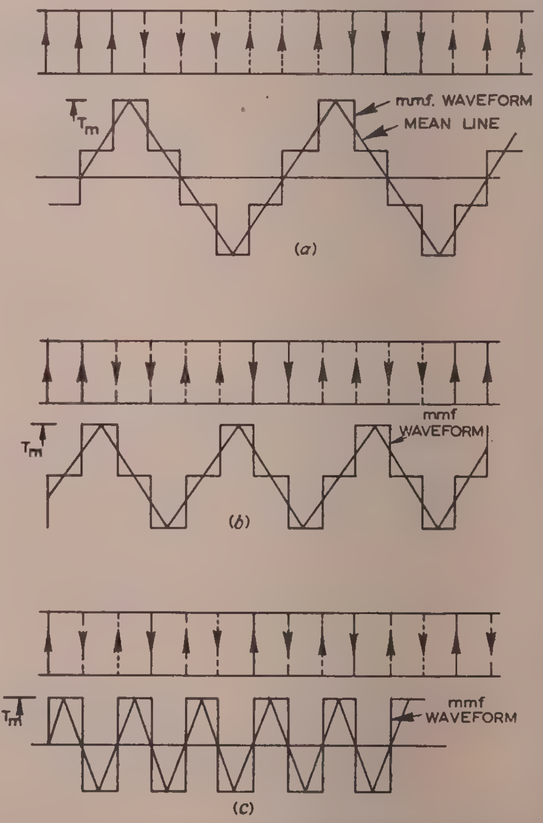


Fig. 3.—Turns distribution of 3-phase full-pitch winding, series-connected.

(a) 3 slots/pole/phase. $v=\frac{8T_m}{3\pi}\left(\sin\theta-\frac{1}{2\times 3}\sin 3\theta+\frac{1}{5}\sin 5\theta+\ldots\right)$

(b) 2 slots/pole/phase. $y=\frac{2\sqrt{2}T_m}{\pi}\left(\sin\theta-\frac{1}{3}\sin 3\theta-\frac{1}{5}\sin 5\theta+\frac{1}{7}\sin 7\theta+\ldots\right)$

(c) 1 slot/pole/phase. $y=\frac{4T_m}{\pi}\left(\sin\theta+\frac{1}{3}\sin 3\theta+\frac{1}{5}\sin 5\theta+\ldots\right)$

ected in star, the star point being brought to a fourth slip ring. he experimental values of V_3/V_1, using each member in turn as he secondary, are shown in Fig. 4 against the theoretical value, vhich runs as a mean line between the experimental curves.

It should be added that r.m.s. meters were used, and that the igher-harmonic content of the voltage V_3, ignored in argument, hus added about 1·5% to the meter reading. On the other and, the meter was known to read about 1% low at 150 c/s, compared with its reading at 50 c/s. These two small errors thus ended to cancel each other, and both were therefore ignored; t was assumed that the experimental values for V_3/V_1 were, in act, values of the ratio of the fundamental quantities, unaffected y harmonics. It will be observed from Fig. 4 that below aturation there is close accord between theory and practice: bove saturation exact analysis would necessarily be exceedingly lifficult.

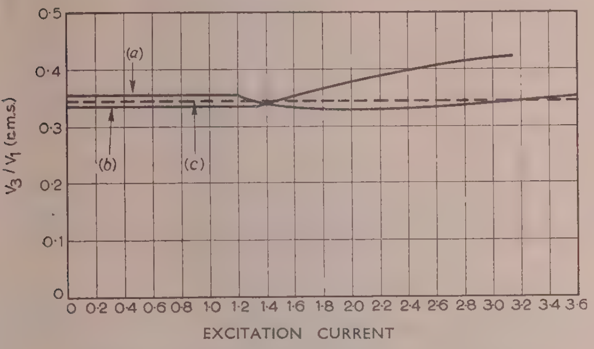


Fig. 4.—Test results for harmonic voltage ratio.
(*a*) Rotor excitation. (*b*) Stator excitation.
(*c*) Theoretical value for particular slotting.

The authors are fully aware that the regulation of a triple-frequency generator based on this principle is very poor, and hey do not suggest that the arrangement is ever likely to be mployed, except experimentally. Load tests have therefore not een recorded, but the close correlation between theory and ractice in regard to open-circuit voltages exhibited by Fig. 4 s worth noting, if only to emphasize the fact that it is easily ttainable even in complicated electrical machine problems, ntil saturation takes control.

5) EFFECT OF SLOTTING THE SECONDARY AND PRIMARY CORES

Preliminary rough tests on a particular machine gave a prac-ical value for V_3/V_1 [eqn. (6)] of about 0·35, taken over the nsaturated range, and it therefore seemed probable that further efinement in calculation was necessary, since this departure from he ideal value of 0·296 exceeded that normally found in tests n unsaturated machines.

The exact forms of m.m.f. distribution, when the secondary vinding was series-connected and fed with direct current, for , 2 and 3 slots/pole/phase, were therefore plotted, the waveforms f turns distribution being shown in Fig. 3; the ratio of the rd-harmonic m.m.f., F_3 (series-excitation), to the fundamental m.m.f., F_1 (excitation as in a synchronous induction motor), is hown in Table 1. It was still assumed that the effects of slotting n the fundamental m.m.f. and spread factor were negligible, .e. the standard results for uniform spread were used for he m.m.f. waveform which is set up by excitation as in a synchronous induction motor. The error in this is, at the worst, ess than 1%, as is well known.

Table 1

CALCULATION OF V_3/V_1, TAKING SLOTTING INTO ACCOUNT

Secondary (d.c.) slots/pole/phase (*s*)	$\dfrac{\text{3rd harmonic m.m.f., } F_3}{\text{Fundamental m.m.f., } F_1}$	Primary 3rd-harmonic spread factor	Primary (a.c.) slots/pole/phase (*q*)
1	$\frac{4}{\pi}(s)\Big/\frac{9}{\pi^2}(2s)=\frac{2\pi}{9}=0\cdot697$	1·000	1
2	$\frac{2\sqrt{2}}{\pi}(s)\Big/\frac{9}{\pi^2}(2s)=\frac{\pi\sqrt{2}}{9}=0\cdot495$	$\frac{1}{\sqrt{2}}=0\cdot707$	2
3	$\frac{8}{3\pi}(s)\Big/\frac{9}{\pi^2}(2s)=\frac{4\pi}{27}=0\cdot465$	$\frac{2}{3}=0\cdot667$	3
∞	$\frac{8}{\pi^2}(s)\Big/\frac{9}{\pi^2}(2s)=\frac{4}{9}=0\cdot444$	$\frac{2}{\pi}=0\cdot637$	∞

*Fundamental spread factor taken as 0·955.
Fundamental m.m.f. F_1 taken as $\frac{9}{\pi^2}$ (central peak m.m.f.).
Both chord factors taken as unity.
For $s=2$ and $q=3$ and for $s=3$ and $q=2$ $\Big\}\ \frac{V_3}{V_1}=\frac{2\sqrt{(2)}\pi^2}{81}=0\cdot345.$

The full Fourier analyses for the waveforms with 1, 2 or 3 slots/pole/phase when series excited are also given in Fig. 3, and these were used for determining the exact waveform to be expected from a particular triple-frequency generator based on these principles. Fig. 3 also shows the triangular waveform of Fig. 2 to which the others tend as the number of slots increases, superimposed in each case.

In addition to allowing for the effect of slotting on the m.m.f. set up by the secondary winding, it was also desirable, for the highest accuracy, to make allowance for the effects of slotting on the spread factor of the primary winding in which the 3rd-harmonic e.m.f. is induced. The values of the 3rd-harmonic spread factor taking account of slotting are given by the exact formula

$$\frac{\sin\left(\frac{m\theta}{2}\right)}{q\sin\left(\frac{m\theta}{2q}\right)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The 3rd-harmonic spread factors in the machine tested were 0·667 and 0·707 for the stator and rotor, respectively, compared with the ideal value for uniform spread of $2/\pi = 0\cdot637$. Again, it was assumed that the effect of slotting on the fundamental induced e.m.f. was negligible, and that the factor for uniform spread (0·955) might be used for both stator and rotor in deducing this.

Taking account of these slot effects, the corrected value of $V_3/V_1 = [2\sqrt{(2)}\pi^2]81 = 0\cdot345$, for the particular machine tested, was obtained (see Table 1), using either the stator or the rotor as secondary winding. This compared with the ideal ratio of 0·296 [eqn. (6)] obtained for uniform slotting.

Repeated careful tests established very close agreement between test and theory when the theoretical ratio appropriate to the slotting was chosen. It is fortunate, for this purpose, that the effect of a limited number of slots is to increase the magnitude of the 3rd harmonic compared with that obtained from a uniform phase spread.

In order that an existing machine shall be used as a 3rd-harmonic generator in the manner discussed, it is necessary that, at least, the star point of both primary and secondary windings shall be available, and preferably that both ends of all three phase windings shall be brought out on either or both members. If the star point only is available on the secondary (d.c.) side, the three phases must be fed in parallel to the star point. If both ends of the phases are available the windings can, more conveniently, be fed in series. If the star point only is available on the primary (a.c.) side, the output must be drawn across each phase to neutral. If both ends are available the primary can be connected either in series or in parallel.

(6) WAVEFORM OF THE TRIPLEN FREQUENCY OUTPUT

The output voltage obtained when the windings are connected for series excitation is basically of three times the fundamental frequency, and includes a series of triplen harmonics which are themselves odd multiples (3, 5, etc.) of three times the fundamental frequency, i.e. of the new basic frequency.

It is easy to deduce a theoretical expression for the resultant waveforms from the m.m.f. analyses given in Fig. 3 and Table 1, together with the appropriate spread factors as calculated from expression (8). It is important to remember that harmonic spread factors can be negative relative to the fundamental,[6] and that only for uniform spread does the numerical magnitude of this factor diminish steadily with increase of the order of harmonic. For a finite number of slots the fundamental spread factor recurs at certain harmonic intervals, the intervals being such that the finite slot "packets" then occur at corresponding points on the particular harmonic flux wave.

The windings of the machine tested were full-pitched to the fundamental, and therefore full-pitched to all odd harmonics as well, and the chording factor could thus be ignored. The chording factor is critical, since a small degree of chording with respect to the fundamental corresponds to a much larger degree of chording of the harmonic. In practice, a machine without virtually full-pitched windings would never be used in the way discussed in this paper. The spread factors here needed are, of course, those for the 3rd, 9th, 15th, 21st, 27th, etc., harmonics of the fundamental, corresponding to the fundamental and to the 3rd, 5th, 7th, 9th, etc., harmonics of the triplen-harmonic-frequency output.

The secondary m.m.f. waveforms, the corresponding primary spread factors and the resultant primary e.m.f.'s using both stator and rotor in turn as secondary, are all given to arbitrary scales in Tables 2 and 3, which form Section 12.2. It will be seen that the ideal resultant e.m.f. waveform is the same whichever limb is excited by direct current. On general principles this must be so, since the e.m.f. induced must always be of the form $i(dM/dt)$, and mutual inductance, M, is ordinarily a reciprocal property, provided that saturation in the magnetic circuit does not occur, or if it occurs, that it is equal whichever of the coils forming the system is used as primary. Clearly then, the ideal equality between voltages will be disturbed in some degree by the effect of saturation of the iron circuit which must always occur to some extent in any actual machine. None the less, this apparently surprising equality was approximately maintained in practice.

There may be some useful practical application of the principle that, when a machine is driven at a given speed, a certain direct current passed through one winding of the machine will give the same induced alternating e.m.f. in the other winding, whichever is chosen for the excitation current. The authors do not know of a practical application, though the principle has here furnished a very useful and interesting check on the accuracy of complicated theoretical analysis.

The theoretical waveform

$$V\left(\sin\theta + \frac{1}{6}\sin 3\theta - \frac{1}{5}\sin 5\theta + \frac{1}{7}\sin 7\theta \ldots, \text{etc.}\right) \quad . \quad (9)$$

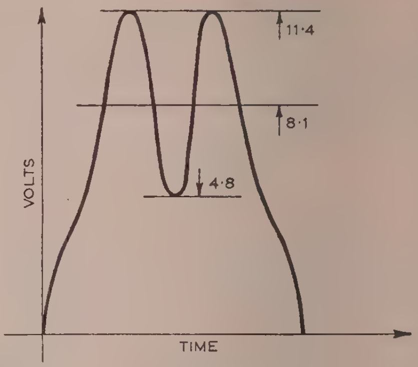


Fig. 5.—Approximate waveform of the series $\sin\theta + \frac{1}{6}\sin 3\theta - \frac{1}{5}\sin 5\theta + \frac{1}{7}\sin 7\theta$ to an arbitrary scale.

is shown to an arbitrary scale, for the first four terms only, in Fig. 5. Preliminary oscillograms of the output waveform showed good general agreement with theory, but the observed magnitude of the harmonics was less than the theoretical. The m.m.f. waveforms shown in Fig. 3, and the assumption that these are the flux waveforms, ignore the effect of fringing, which will become more marked for the higher harmonics.

Tests of the harmonic content of the actual generated e.m.f.'s were accordingly made with a harmonic analyser, and from these values a curve of output voltage was constructed which closely agreed with the oscillogram recording. The actual waveform was similar for both stator and rotor excitation, but, for the reason discussed above, the results were not identical. The approximate waveforms obtained were

$$A(\sin\theta + 0{\cdot}15\sin 3\theta - 0{\cdot}20\sin 5\theta + 0{\cdot}04\sin 7\theta) \quad (10)$$

for stator excitation, and

$$A(\sin\theta + 0{\cdot}20\sin 3\theta - 0{\cdot}08\sin 5\theta + 0{\cdot}03\sin 7\theta) \quad (11)$$

for rotor excitation, where A is an arbitrary constant.

For harmonics higher than the 7th the experimental values of amplitude diminished very much more rapidly than the theoretical values given by expression (9), and they were virtually negligible.

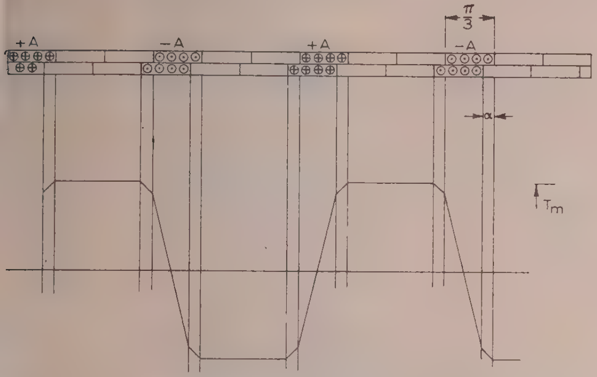
(7) EFFECT OF CHORDING THE WINDINGS

If either the primary winding in which the voltages are induced, or the secondary winding which carries the series-excitation current, are chorded, a considerable variation in the 3rd-harmonic voltage will result.

The effect of chording the secondary winding will first be considered. The turns distribution of a single phase of a winding of 60° spread, chorded back by an angle α, is shown in Fig. 6. Its Fourier analysis can be shown, by the usual methods, to be

$$\frac{12}{\pi^2}T_m\left(\sin\theta\cos\frac{\alpha}{2} + \frac{2}{9}\sin 3\theta\cos\frac{3\alpha}{2} + \frac{1}{25}\sin 5\theta\cos\frac{5\alpha}{2}\right.$$
$$-\frac{1}{49}\sin 7\theta\cos\frac{7\alpha}{2} - \frac{2}{81}\sin 9\theta\cos\frac{9\alpha}{2}$$
$$\left.-\frac{1}{121}\sin 11\theta\cos\frac{11\alpha}{2} + \ldots\right) \quad . \quad (12)$$

This expression is identical with expression (3), except that each term is multiplied by $\cos(m\alpha/2)$. Now it is already known that only odd multiples of the third harmonic will appear in the



ig. 6.—Turns distribution of one phase of a 60°-spread winding, chorded by an angle α.

esultant m.m.f. owing to series-excitation of the three phases; will be seen that the magnitudes of these harmonics will be educed in the ratios $\cos(3\alpha/2)$, $\cos(9\alpha/2)$, $\cos(15\alpha/2)$, etc., nd the Fourier analysis of the resultant m.m.f. of the three hases, when series-excited by a current I_m, is seen at once, sing the arguments of Section 2, to be

$$\frac{3}{2}I_mT_m\left(\sin 3\theta\cos\frac{3\alpha}{2}-\frac{1}{3^2}\sin 9\theta\cos\frac{9\alpha}{2}\right.$$
$$\left.+\frac{1}{5^2}\sin 15\theta\cos\frac{15\alpha}{2}-\ldots\right) . \quad (13)$$

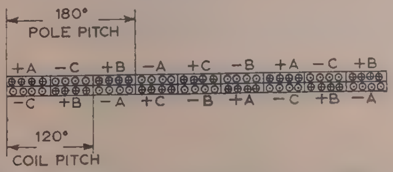
y Section 2 also, the resultant m.m.f. of the three phases when xcited as in a synchronous induction motor, with a total direct urrent I_m, is

$$I_mT_m\frac{9}{\pi^2}\left(\sin\theta\cos\frac{\alpha}{2}+\frac{1}{5^2}\sin 5\theta\cos\frac{5\alpha}{2}\right.$$
$$\left.-\frac{1}{7^2}\sin 7\theta\cos\frac{7\alpha}{2}-\frac{1}{11^2}\sin 11\theta\cos\frac{11\alpha}{2}+\ldots\right) . \quad (14)$$

his result was also obtained in different form by direct analysis y Jakeman,[7] in a paper which should be consulted in relation o the effects of chording on m.m.f. curves. The result given in xpression (13), for the m.m.f. wave actually obtained by series xcitation of the three chorded phase windings, can otherwise e obtained by direct analysis. Because the result requires care n its direct deduction and is not exactly what might be expected t first sight, a brief outline of the proof is given in Section 12.1.

From an inspection of expression (13) it is apparent that as ne chording is increased the higher terms will very rapidly iminish to zero, and indeed pass through it.

In the limit, where $\alpha = \pi/3$ (that is, for 2/3 chording) all the hord factors, $\cos(3\alpha/2)$, $\cos(9\alpha/2)$, etc., will be simultaneously ero, and the combined m.m.f. of all the windings when series-xcited will be zero. This is of course clear by reference to the hase-band distribution diagram (Fig. 7) for a double-layer winding of 60° spread, chorded to two-thirds (120°) of full pitch. The upper layer is, at every point, neutralized by the lower layer, and the three windings together produce no flux.



ig. 7.—Phase-band diagram of double-layer 60°-spread, 120°-coil-pitch, series-fed winding, showing total neutralization of all m.m.f.'s.

If, however, the secondary winding is full-pitched, so that it sets up an m.m.f., the induced e.m.f.'s in the primary winding can still be reduced or eliminated by chording the primary winding in the usual way. In particular, two-thirds chording of the primary winding will eliminate the 3rd-harmonic induced e.m.f., and its family of triplen harmonics.

It may well be that the results given in this Section are of less significance than the rest of the paper, but it has been thought desirable to include them for completeness, since they might well be applicable in other contexts.

(8) MAGNETIZING REACTANCES WITH VARIOUS TYPES OF WINDING

(8.1) General Discussion

It is possible, as explained in a companion paper,[8] to reconnect the winding of a given machine for a 3 : 1 pole and speed ratio in several different ways besides the prototype discussed above. The theoretical ratios (x_1/x_3) between the two magnetizing reactances per phase, for each of several types of pole-changing connection, are deduced below, and certain experimental results are added.

The ratio is deduced for each of the following çases:

(*a*) When first connected 3-phase delta ($2p$ poles), and then connected single-phase open-delta ($6p$ poles). This gives pole-changing without any alteration in the windings of each phase and is the type of connection described in this paper.

This ratio can otherwise be described as the ratio between the positive-sequence magnetizing reactance x_1 and the zero-sequence magnetizing reactance x_3, per phase, of the given winding.

(*b*) When first connected 3-phase delta ($2p$ poles), and then reconnected 3-phase star ($6p$ poles). This was the form first tested by the authors, though previously suggested by others. It is discussed in the companion paper.[8]

(*c*) When first connected 3-phase delta, two-thirds full spread ($2p$ poles), and then reconnected 3-phase star ($6p$ poles). This was the second form tested by the authors and found successful, apart from crawling torques for the $2p$-poles connection.

All these variations were first considered, both in theory and in practice, for a double-layer winding connected in the standard industrial manner for 60° spread. When at a later stage in the tests it was decided to reconnect the winding as a true 3-phase winding for 120° spread, it seemed desirable to repeat these computations, and the confirmatory experiments, for 120° spread.

It is not, however, possible to obtain any results for the open-delta connection with a winding of 120° spread, because the reactance of such a winding, series-fed in open delta, is zero: the winding generates no flux and the machine ceases to function as a piece of dynamo-electric apparatus. This will be clear at once from Fig. 8, which shows that a 120°-spread winding gives zero net m.m.f. in every slot. The point may be expressed otherwise by saying that the m.m.f. of a single-phase winding of 120° spread has no triplen harmonic content, and that the resultant

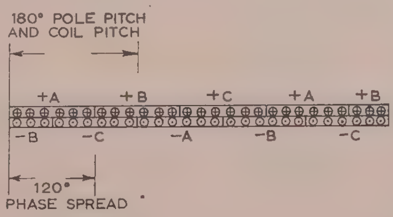


Fig. 8.—Phase-band diagram of double-layer 120°-spread full-pitch, series-fed winding, showing total neutralization of all m.m.f.'s.

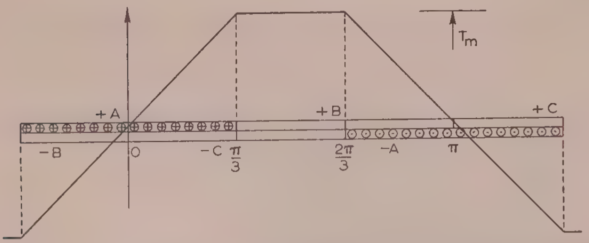


Fig. 9.—Turns distribution of one phase of a 120°-spread full-pitch winding.

of the three phase-windings in series is necessarily zero. The Fourier analysis of a single phase of 120° spread, as shown in Fig. 9, can readily be obtained as

$$\frac{6\sqrt{3}}{\pi^2}T_m\left(\sin\theta - \frac{1}{5^2}\sin 5\theta + \frac{1}{7^2}\sin 7\theta - \frac{1}{11^2}\sin 11\theta + \ldots, \text{etc.}\right)$$

This result should be contrasted with the result given by Fig. 1 and by expression (3), for a single phase of 60° spread.

In all these calculations it is assumed to be sufficient to compute the fundamental space-component of m.m.f., and to assume that this alone sets up a flux which induces voltage. The voltage induced by the space-harmonic fluxes will be exceedingly small. It is further assumed that the full terminal voltage is applied to a totally inductive magnetizing circuit. In fact, there is always an equivalent series inductance—the primary leakage inductance—through which the magnetizing current must also flow. The relative magnitude of this inductance is very much greater for the larger number of poles, and, in consequence, the practical value of the ratio x_3/x_1 always exceeds the theoretical value. The leakage reactance of the winding in any connection can, of course, be determined easily by the usual short-circuit test on low voltage. This was done in a number of cases, and whilst the detailed results have not been thought worth inclusion, it may be recorded that the measured ratios of leakage reactance to magnetizing reactance were of the order which would be required to account for the practical value of the ratio x_3/x_1 exceeding the theoretical ratio by amounts varying from 15 to 34%, which is the range of variation shown in Table 2.

Table 2

RATIOS OF MAGNETIZING REACTANCES FOR VARIOUS CONNECTIONS

Connection	Value of $\frac{x_3}{x_1}$		Ratio $\left(\frac{\text{Practical}}{\text{Theoretical}}\right)$
	Theoretical	Practical	
60° spread, 3-phase delta, and single-phase open delta.	0·103	0·119	1·15
60° spread, 3-phase delta, and 3-phase star, reconnected.	0·122	0·163	1·34
60° spread, 3-phase delta ($\frac{2}{3}$ used), and 3-phase star, reconnected.	0·260	0·331	1·27
120° spread, 3-phase delta, and 3-phase star, reconnected.	0·122	0·144	1·18
120° spread, 3-phase delta ($\frac{2}{3}$ used), and 3-phase star, reconnected.	0·225	0·282	1·25

The practical values of the ratio x_3/x_1 were deduced in a typical case from the initial slopes of the families of magnetizing curves which are shown in Figs. 14 and 15, and which refer to the machine when connected respectively with 60° spread and with 120° spread; Fig. 15, therefore, shows no magnetizing curve for open-delta operation.

In Table 2 is given a summary of the values of the ratios of x_3/x_1 for the three types of connection as obtained by calculation and also by experiment. The value 0·103, obtained with the first connection given, is the fixed ratio of the zero-sequence magnetizing reactance to the positive-sequence magnetizing reactance for a 60°-spread winding, with 3 slots/pole/phase. This ratio is independent of the number of conductors in the winding, but would vary somewhat if the number of slots per pole per phase were varied.

Calculation of the theoretical ratios is given in Sections 8.2 to 8.6.

(8.2) 3-Phase Delta (2*p* poles), and Single-Phase Open-Delta (6*p* poles), both 60° Spread

For equal flux density in the two cases, which is the condition for equating magnetizing force, it follows that

$$\frac{\phi_3}{\phi_1} = \frac{1}{3}$$

Now $\dfrac{V_3}{V_1} = \dfrac{\phi_3}{\phi_1} \times \dfrac{\text{Spread factor for higher pole number}}{\text{Spread factor for lower pole number}}$

since the total number of conductors per phase is the same in both cases. On the assumption that the winding may be regarded as uniformly spread in both cases,

$$\frac{V_3}{V_1} = \frac{\phi_3}{\phi_1}\frac{2/\pi}{3/\pi} = \frac{2}{9} = 0{\cdot}222 \quad . \quad . \quad . \quad . \quad (15)$$

For the higher number of poles the phases are in arithmetic series and therefore

$$\text{Open delta voltage/Line voltage in delta} = \frac{2}{3} = 0{\cdot}667$$

For equal flux density, the magnetizing force acting on the centre of the resultant poles must be the same for both numbers of poles.

The peak fundamental m.m.f. in ampere-turns per pole is, as is well known

$$F_1 = \frac{9\sqrt{2}}{\pi^2}(nI_1) = 1{\cdot}29nI_1$$

By reference to Fig. 3, and considering the case of 3 slots per pole per phase in the original winding, it will be clear that the peak m.m.f. is given, for single-phase open-delta operation, by

$$F_3 = \frac{8}{3\pi}\frac{n}{2}\sqrt{(2)}I_3 = \frac{4\sqrt{2}}{3\pi}(nI_3) = 0{\cdot}60nI_3$$

Equating F_1 and F_3,

$$\frac{I_3}{I_1} = \frac{27}{4\pi} = 2{\cdot}15 \quad . \quad . \quad . \quad . \quad . \quad (16)$$

Combining this with the value already deduced for V_3/V_1, we obtain the ratio of magnetizing reactances

$$\frac{x_3}{x_1} = \frac{\frac{V_3}{V_1}}{\frac{I_3}{I_1}} = \frac{8\pi}{243} = 0{\cdot}103 \quad . \quad . \quad . \quad . \quad (17)$$

This is the ideal ratio of the magnetizing reactances per phase, when a delta-connected 3-phase motor is operated first normally and then single phase in open-delta, with three times the number of poles and at one-third the speed. It is also the ratio of positive-sequence magnetizing reactance to zero-sequence magnetizing reactance for a normal 3-phase winding having 3 slots/pole/phase.

(8.3) 3-Phase Delta (2*p* poles), and Reconnected 3-Phase Star (6*p* poles), both 60° Spread

It can readily be seen that in this case

$$\frac{V_3}{V_1} = \frac{1}{3}\,\frac{1}{3/\pi} = \frac{\pi}{9} \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

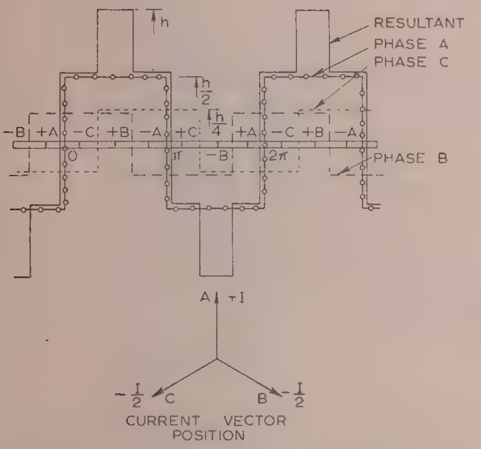


Fig. 10.—M.M.F. waveform of 60°-spread winding with one slot per pole per phase.

$h = \sqrt{2}nI_a$ ampere-turns/pole.

Fundamental Fourier component = Peak value $\times \frac{3}{\pi}$.

assuming 60° uniform spread in the first case and a winding concentrated in one slot per pole per phase in the second case.

The peak fundamental m.m.f. in the first case is, as before,

$$F_1 = \frac{9\sqrt{2}}{\pi^2}nI_1 = 1{\cdot}29nI_1$$

By reference to Fig. 10, and to the Fourier analysis of the stepped wave which it shows for the larger number of poles, it will be clear that the peak fundamental m.m.f. in the reconnected case, if there are initially 3 slots/pole/phase, is

$$F_3 = \frac{3\sqrt{2}}{\pi}\,\frac{nI_3}{3} = \frac{\sqrt{2}}{\pi}nI_3 = 0{\cdot}45nI_3$$

where the number of conductors per pole per phase is now *n*/3. If the change of waveform is ignored, and the standard 3-phase 60° spread m.m.f. waveforms are still assumed, the result would be $F_3 = \frac{9\sqrt{2}}{\pi^2}\,\frac{nI_3}{3} = 0{\cdot}43nI_3$ which is only slightly different numerically.

Equating F_1 and F_3

$$\frac{I_3}{I_1} = \frac{9}{\pi} = 2{\cdot}86 \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Thus

$$\frac{x_3}{x_1} = \frac{\dfrac{V_3}{V_1}}{\dfrac{I_3}{I_1}} = \frac{\pi^2}{81} = 0{\cdot}122 \quad . \quad . \quad . \quad . \quad (20)$$

This is the ratio of magnetizing reactances per phase for the prototype pole-changing winding with 60° spread in its two connections. It is a distinct improvement on the value 0·103 for an open-delta connection.

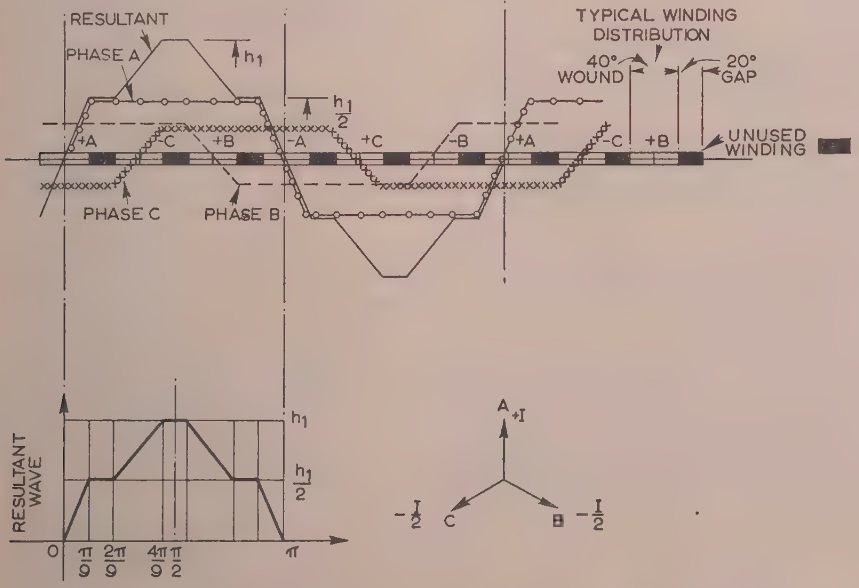


Fig. 11.—M.M.F. waveform of 60°-spread full-pitch double-layer winding (only 40° spread used, followed by 20° gap), for currents as shown by vectors.

Vector position I. $h_1 = \frac{2}{3}\sqrt{2}I_a n$ ampere-turns/pole. Coefficient of *m*th harmonic is $a_m = \frac{36h_1}{\pi^2 m^2}\left(\cos^2\frac{m\pi}{6}\sin\frac{m\pi}{9}\right)$

(8.4) 3-Phase Delta, Two-thirds (40°) Spread (2*p* poles), and Reconnected 3-Phase Star 60° Spread (6*p* poles)

By the same reasoning as before, we now have:

$$\frac{\phi_1}{\phi_3} = 3 \frac{\dfrac{V_1}{\frac{2}{3}n \times 0\cdot979}}{\dfrac{V_3}{n \times 1\cdot000}} = \frac{V_1}{V_3} \times \frac{3}{2} \times \frac{1\cdot000}{0\cdot979}$$

or

$$\frac{V_1}{V_3} = 1\cdot96 \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

Only $2n/3$ conductors per phase are used with $2p$ poles, and the factor 0·979 is the spread factor for a 40° uniformly spread phase band. The winding for $6p$ poles is concentrated in one slot per pole per phase.

Taking the m.m.f. waveform from Fig. 11, the m.m.f. for $2p$ poles can be written

$$F_1 = 0\cdot935 \frac{2}{3} n\sqrt{(2)}I_1 = 0\cdot883nI_1$$

For $6p$ poles, the expression for waveform is the same as in Section 8.3, i.e.

$$F_3 = \frac{3\sqrt{2}}{\pi} \frac{n}{3} I_3 = 0\cdot45nI_3$$

Equating F_1 and F_3,

$$\frac{I_3}{I_1} = 0\cdot935 \frac{2\pi}{3} = 1\cdot96 \quad . \quad . \quad . \quad . \quad (22)$$

The ratio of magnetizing reactances then follows as

$$\frac{x_3}{x_1} = \frac{\dfrac{V_3}{V_1}}{\dfrac{I_3}{I_1}} = 0\cdot260 \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

It will be observed that the effect of omitting one-third of the winding with $2p$ poles is desirably to increase the ratio x_3/x_1 to more than twice its previous value of 0·122.

(8.5) 3-Phase Delta (2*p* poles), and Reconnected 3-Phase Star (6*p* poles), both 120° Spread

For the same reasons as for the connection discussed in Section 8.3, it follows that

$$\frac{V_3}{V_1} = \frac{1}{3} \frac{\left(\dfrac{\sqrt{3}}{2}\right)}{\left(\dfrac{3\sqrt{3}}{2\pi}\right)} = \frac{\pi}{9} = 0\cdot349 \quad . \quad . \quad . \quad (24)$$

the spread factor now being $3\sqrt{(3)}/2\pi$ for $2p$ poles, corresponding to 120° uniform spread, and $\sqrt{(3)}/2$ when reconnected, corresponding to 120° spread with two slots per 120° phase band. It should be noted that with 120° spread the number of slots per phase band is twice the number of slots per pole per phase.

The peak fundamental m.m.f. for $2p$ poles, as is well known, is now

$$F_1 = \frac{9}{\pi^2} \frac{\sqrt{3}}{\sqrt{2}} nI_1 = 1\cdot12nI_1$$

the effect of using 120° spread being to reduce the magnetizing force in the ratio $\sqrt{(3)}/2$, compared with 60° spread.

By reference to Fig. 12 it will be clear that the peak fundamental m.m.f. in the reconnected condition (if there are initially 3 slots/pole/phase, and therefore only one slot per pole per phase on reconnection) is

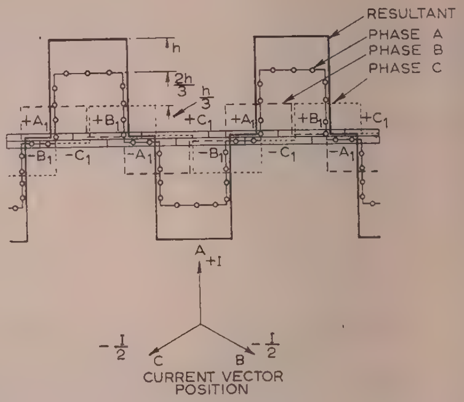


Fig. 12.—M.M.F. waveform of 120°-spread winding with one slot per pole per phase.

$h = \dfrac{3}{2\sqrt{2}} I_a n$ ampere-turns/pole.

Fundamental Fourier component = Peak value × $\dfrac{2\sqrt{3}}{\pi}$.

1 slot/pole/phase = 2 slots/phase-band with 120°-spread double-layer winding.

$$F_3 = \frac{3\sqrt{3}}{\pi\sqrt{2}} \frac{nI_3}{3} = 0\cdot39nI_3$$

where the number of slots per pole per phase is now $n/3$. If the change of waveform is ignored and the standard 3-phase 120°-spread m.m.f. waveforms are still assumed, the result would be

$$F_3 = \frac{9}{\pi^2} \frac{\sqrt{3}}{\sqrt{2}} \frac{nI_3}{3} = 0\cdot372nI_3$$

This again differs numerically, but not appreciably, from the more exact value, as it did for 60° spread.

Equating F_3 and F_1,

$$\frac{I_3}{I_1} = \frac{9}{\pi} = 2\cdot86 \quad . \quad . \quad . \quad . \quad . \quad (25)$$

from which

$$\frac{x_3}{x_1} = \frac{\dfrac{V_3}{V_1}}{\dfrac{I_3}{I_1}} = \frac{\pi^2}{81} = 0\cdot122 \quad . \quad . \quad . \quad . \quad (26)$$

It will thus be seen that the ratio of magnetizing reactances per phase, for the prototype pole-changing winding but with 120° spread, is the same as the ratio with 60° spread.

(8.6) 3-Phase Delta, Two-thirds (80°) Spread (2*p* poles), and Reconnected 3-Phase Star 120° Spread (6*p* poles)

For the same reasons as in the case discussed in Section 8.4, it follows that for this winding,

$$\frac{\phi_1}{\phi_3} = 3 = \frac{\dfrac{V_1}{\frac{2}{3}n \times 0\cdot923}}{\dfrac{V_3}{n \times 0\cdot866}} = \frac{V_1}{V_3} \times \frac{3}{2} \times \frac{0\cdot866}{0\cdot923}$$

or

$$\frac{V_1}{V_3} = 2\cdot13 \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

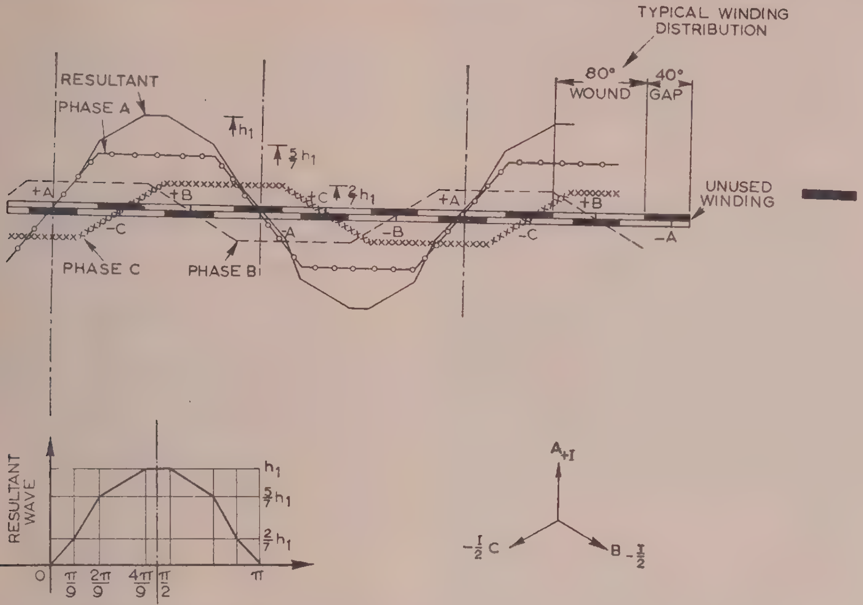


Fig. 13.—M.M.F. waveform of 120°-spread full-pitch double-layer winding (only 80° spread used, followed by 40° gap) for currents as shown by vectors.

$h_1 = \frac{7}{6\sqrt{2}} I_a n$ ampere-turns/pole. Coefficient of mth harmonic is $a_m = \frac{144h_1}{7\pi^2 m^2}\left(\cos^2 \frac{m\pi}{6} \sin \frac{2m\pi}{9}\right)$

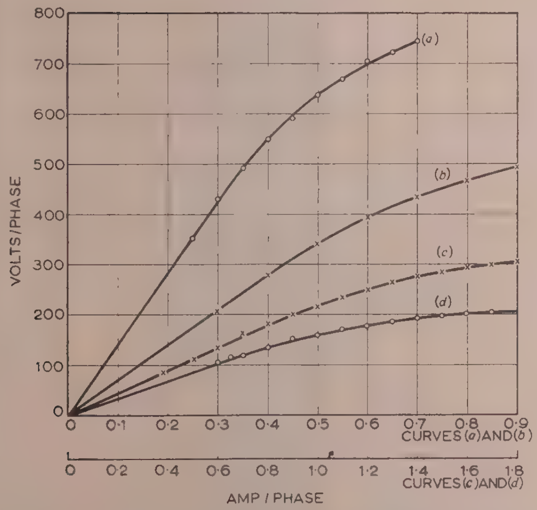


Fig. 14.—Magnetizing curves for typical 4/12-pole induction motor, with 60°-spread full-pitch windings, for various connections.

(a) Normal delta, 4 poles. x_1 = 1 380 ohms.
(b) Delta 40°-spread, 4 poles. x_1 = 680 ohms.
(c) Star reconnected, 12 poles. x_3 = 225 ohms.
(d) Open delta single-phase, 12 poles. x_3 = 165 ohms.

Only (2/3)n conductors per phase are used with $2p$ poles; the factor 0·923 is the spread factor for an 80° uniformly spread phase band, and 0·866 is the same spread factor as for the case discussed in Section 8.5 with $6p$ poles.

Taking the m.m.f. waveform from Fig. 13, the m.m.f. in the $2p$-poles connection can be written as

$$F_1 = 1\cdot00 \frac{7}{6\sqrt{2}} nI_1 = 0\cdot823nI_1$$

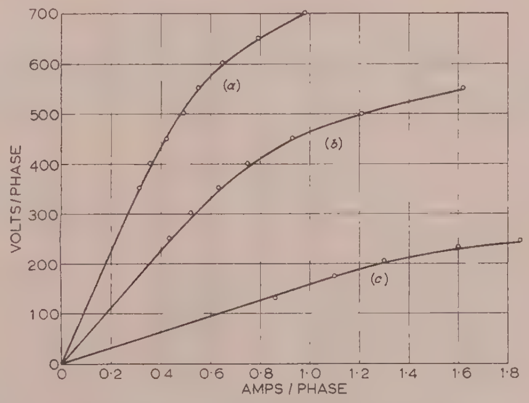


Fig. 15.—Magnetizing curves for typical 4/12-pole induction motor with 120°-spread full-pitch windings, for various connections.

(a) Normal delta, 4 poles. x_1 = 1 110 ohms.
(b) Delta 80°-spread, 4 poles. x_1 = 565 ohms.
(c) Star reconnected, 12 poles. x_1 = 160 ohms.

For $6p$ poles, the expression is the same as in Section 8.5 for $6p$ poles, i.e.

$$F_3 = \frac{\sqrt{3}}{\pi\sqrt{2}} nI_3 = 0\cdot390nI_3$$

Equating F_1 and F_3,

$$\frac{I_3}{I_1} = \frac{1\cdot00 \times 7\pi}{6\sqrt{3}} = 2\cdot12 \quad . \quad . \quad . \quad . \quad (28)$$

and therefore

$$\frac{x_3}{x_1} = \frac{\frac{V_3}{V_1}}{\frac{I_3}{I_1}} = 0\cdot222 \quad . \quad . \quad . \quad . \quad . \quad (29)$$

This is a considerable improvement on the value 0·122 when using the whole winding, although it is not quite so favourable, relatively, as the value 0·260 for 60° spread using two-thirds of the winding. The higher value arises from the concentration of the winding which for the 6*p*-poles connection discussed in Section 8.4 is in one slot, while in this case it is spread over two slots per phase band.

(9) CONCLUSION

The generator arrangement described here is of some theoretical interest because it exemplifies and uses the zero-sequence field, and this extension of the method of "freezing" a.c. vectors (hitherto used only in the synchronous induction motor and in various kinds of Magslip) to zero-sequence currents is of special analytical interest. In practice, this type of generator may sometimes be used to improvise a triple-frequency supply from standard equipment for test purposes, but a 3 : 1 3-phase pole-changing winding, developed from the same theory, is very much more likely to be industrially useful. This winding has properties superior to most commonly accepted pole-changing windings, and has accordingly been investigated in detail and described in a separate paper.[8]

The other results given in the paper are also primarily of theoretical interest, but recent growing concern with asymmetrical windings and connections has given a new value to experiment and analysis on the zero-sequence components of unbalanced systems. The authors believe that a too slavish addiction to balanced positive-sequence systems only has perhaps tended to limit the vision of some machine designers, and that the effect of zero-sequence phenomena has sometimes been overlooked.

(10) ACKNOWLEDGMENTS

The authors wish to acknowledge the help given by Dr. J. E. Brown of the University of Bristol. They are also indebted to Mr. I. R. Smith and Mr. B. V. Jayawant who assisted in the verification of theoretical calculations and practical observations.

(11) REFERENCES

(1) SMITH, S. P., and BOULDING, R. S. H.: "The Shape of the Pressure Wave in Electrical Machinery," *Journal I.E.E.*, 1915, **53**, p. 205.
(2) HAGUE, B.: "The Mathematical Treatment of the M.M.F. of Armature Windings," *ibid.*, 1917, **55**, p. 489.
(3) RAWCLIFFE, G. H.: "The Secondary Circuits of Synchronous Induction Motors," *ibid.*, 1940, **87**, p. 282.
(4) BROWN, J. E., and BUTLER, O. I.: "A General Method of Analysis of Three-phase Induction Motors with Asymmetrical Primary Connections," *Proceedings I.E.E.*, Paper No. 1421 U, February, 1953 (**100**, Part II, p. 25).
(5) CLAYTON, A. E.: "A Mathematical Development of the Theory of the Magnetomotive Force of Windings," *Journal I.E.E.*, 1923, **61**, p. 749.
(6) LIWSCHITZ-GARIK, M., and WHIPPLE, G. C.: "Electrical Machinery," Vol. II (Van Nostrand), pp. 129–135.
(7) JAKEMAN, R. G.: "Analysis of the M.M.F. Curves of Short-Chorded Windings," *G.E.C. Journal*, 1940, **11**, p. 66.
(8) RAWCLIFFE, G. H., and JAYAWANT, B. V.: "The Development of a New Three-to-One Pole-Changing Motor," *Proceedings I.E.E.*, Paper No. 1958 U, December, 1955 (**103 A**).

(12) APPENDICES

(12.1) Fourier Analysis for Series-Excitation of Short-Chorded Windings

As the windings are progressively chorded, the upper and lower layers of the winding begin to neutralize each other.

Table 3.—DEDUCTION OF WAVEFORM OF TRIPLE-FREQUENCY GENERATOR

	Stator excited	Rotor excited
Secondary slots/pole/phase, *s*	3	2
Secondary m.m.f. for series connection with unit current on 3rd-harmonic scale of θ. Fourier series as in Fig. 3	$y = \frac{8T_m}{3\pi}\Big(\sin\theta - \frac{1}{2\times 3}\sin 3\theta + \frac{1}{5}\sin 5\theta + \frac{1}{7}\sin 7\theta$ $- \frac{1}{2\times 9}\sin 9\theta + \frac{1}{11}\sin 11\theta + \frac{1}{13}\sin 13\theta - \frac{1}{2\times 15}\sin 15\theta$ $+ \frac{1}{17}\sin 17\theta + \frac{1}{19}\sin 19\theta - \frac{1}{2\times 21}\sin 21\theta + \frac{1}{23}\sin 23\theta + \ldots\Big)$	$y = \frac{2\sqrt{2}T_m}{\pi}\Big(\sin\theta - \frac{1}{3}\sin 3\theta - \frac{1}{5}\sin 5\theta + \frac{1}{7}\sin 7\theta + \frac{1}{9}\sin 9\theta$ $- \frac{1}{11}\sin 11\theta - \frac{1}{13}\sin 13\theta + \frac{1}{15}\sin 15\theta + \frac{1}{17}\sin 17\theta$ $- \frac{1}{19}\sin 19\theta - \frac{1}{21}\sin 21\theta + \frac{1}{23}\sin 23\theta + \ldots\Big)$
Primary slots/pole/phase, *q*	2	3
Primary spread factors for corresponding harmonics, from eqn. (8)	$+\frac{1}{\sqrt{2}};\ -\frac{1}{\sqrt{2}};\ -\frac{1}{\sqrt{2}};\ +\frac{1}{\sqrt{2}};\ +\frac{1}{\sqrt{2}};\ -\frac{1}{\sqrt{2}}$ $-\frac{1}{\sqrt{2}};\ +\frac{1}{\sqrt{2}};\ +\frac{1}{\sqrt{2}};\ -\frac{1}{\sqrt{2}};\ -\frac{1}{\sqrt{2}};\ +\frac{1}{\sqrt{2}}$	$+\frac{2}{3};\ -\frac{1}{3};\ +\frac{2}{3};\ +\frac{2}{3};\ -\frac{1}{3};\ +\frac{2}{3};$ $+\frac{2}{3};\ -\frac{1}{3};\ +\frac{2}{3};\ +\frac{2}{3};\ -\frac{1}{3};\ +\frac{2}{3}$
Fourier series of e.m.f. corresponding to Fig. 5	$y \propto \frac{4\sqrt{2}T_m}{3\pi}\Big(\sin\theta + \frac{1}{2\times 3}\sin 3\theta - \frac{1}{5}\sin 5\theta + \frac{1}{7}\sin 7\theta - \frac{1}{2\times 9}\sin 9\theta - \frac{1}{11}\sin 11\theta - \frac{1}{13}\sin 13\theta$ $- \frac{1}{2\times 15}\sin 15\theta + \frac{1}{17}\sin 17\theta - \frac{1}{19}\sin 19\theta + \frac{1}{2\times 21}\sin 21\theta + \frac{1}{23}\sin 23\theta + \ldots\Big)$	

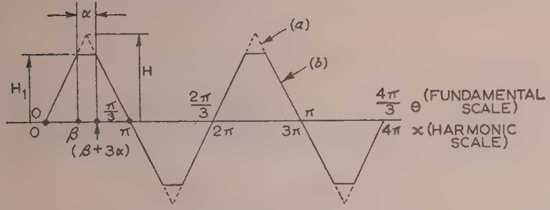


Fig. 16.—M.M.F. waveform due to series-fed, 60°-spread, 3-phase chorded windings.

$$H_1 = H\left(\frac{\pi - 3\alpha}{\pi}\right).\quad \beta = \frac{\pi}{2} - \frac{3\alpha}{2}.\quad x = 3\theta.$$

(*a*) Triangular wave: resultant m.m.f. of three unchorded phases.
(*b*) Trapezoidal wave due to chording.

The effect is to remove the peaks of the initial triangular m.m.f. wave in the manner shown in Fig. 16 and to set up a trapezoidal m.m.f. wave of amplitude diminishing as the chording increases. In the limit, where $\alpha = \frac{\pi}{3}$, the m.m.f. wave vanishes at all points.

The Fourier analysis of this wave on the natural scale is

$$\frac{4H_1}{\pi\beta}\left(\sin\beta\sin x + \frac{1}{3^2}\sin 3\beta\sin 3x + \frac{1}{5^2}\sin 5\beta\sin 5x + \ldots\right)$$

Rewrite in terms of α, *H* and θ, using the relationships shown below the Figure, and the analysis then is

$$\frac{8H}{\pi^2}\left(\cos\frac{3\alpha}{2}\sin 3\theta - \frac{1}{3^2}\cos\frac{9\alpha}{2}\sin 9\theta + \frac{1}{5^2}\cos\frac{15\alpha}{2}\sin 15\theta - \ldots\right)$$

which is identical with expression (13).

(12.2) Theoretical Expressions for Induced E.M.F.'s for Series-Excited Machine

As has been explained in Section 6, the 3rd-harmonic induced e.m.f. for a given exciting current must in principle be the same whether the stator or the rotor is excited. It is, however, instructive to carry out the calculation in detail and to verify that the result is identical for both cases. This is done in Table 3. The higher harmonics are small, but it is necessary to take 12 terms in order to pass through a whole cycle of signs.

621.385.032.3.017.7

Monograph No. 158 R
Dec. 1955

MUTUAL HEATING IN TRANSMITTING-VALVE FILAMENT STRUCTURES

By W. J. POHL, M.Sc., Associate Member.

(*The paper was first received* 17*th June, and in revised form* 25*th August*, 1955. *It was published as an* INSTITUTION MONOGRAPH *in December*, 1955.)

SUMMARY

The paper deals with heat radiation characteristics of filament structures in the region of 2000° K, such as are used in transmitting valves. It shows how the effects of mutual heating between the individual elements may be calculated, and includes a set of universal curves which enable the results to be readily applied to structures of cylindrical form.

(1) INTRODUCTION

In the design of filament structures for transmitting valves, the highly critical nature of the effect of temperature on valve life and on emission is widely recognized.[1] In the future, an increasing requirement for valves operating at higher frequencies will necessitate filament structures with large numbers of wires to give large emitting areas, and to minimize the self-inductance of the filament. For example, a modern high-frequency transmitting triode for operation at frequencies up to 250 Mc/s may have a filament structure consisting of 50 or more thin thoriated tungsten strands in parallel. In such structures the mutual heating effect can be appreciable, so that normal design methods applicable to straight single wires[1,2,3] lead to higher values of filament power than are necessary to attain required levels of emission. The paper shows how to calculate the extent to which mutual heating gives a saving in power at any stipulated temperature. The results of the work it describes are given in Sections 2, 3 and 7, and these form a self-contained Summary.

(2) THE FACTOR K

The power radiated per unit length from a straight rod or wire of circular or rectangular section is proportional to A, the surface area per unit length. If the wire is at a uniform temperature the power can be expressed as

$$\text{Radiated power} = \sigma\phi A(T^4 - T_0^4)$$

where σ is the Stefan–Boltzmann constant, T and T_0 are the temperatures of the wire and the surroundings respectively in degrees absolute and ϕ is the emissivity. This is the Stefan–Boltzmann law and applies when the wire is enclosed by a perfect black body. The figures given for tungsten wire in References 2 and 3 apply *in vacuo* under these conditions. Since $T \gg T_0$ the expression can be written

$$\text{Radiated power} = \sigma\phi AT^4 \quad . \quad . \quad . \quad . \quad (1)$$

If some of the heat is reflected, however, or if the wire receives heat from another wire, the conditions may be represented by multiplying the emissivity by a factor K less than unity, and the equation will then be

$$\text{Radiated power} = K\sigma\phi AT^4$$

It is proposed to calculate the value of K in cylindrical filament structures in which the elements are arranged in a symmetrical manner, as shown in Fig. 1, although the results are applicable to any cylindrical arrangement of wires or tapes such as, say, a helical one. End-lead conduction will be neglected since, at operating temperatures around 2000° K, its effects are usually confined to a short length at the end of each wire, and can be calculated by the use of curves given in Reference 4. Only that portion of the wire over which conduction cooling causes no variation in temperature is considered, and this is usually referred to as the "effective emitting length." This term will be used throughout the paper.

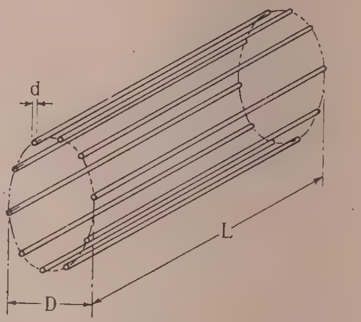


Fig. 1.—Typical filament structure.

If I_1 is the calculated current required for a short section of wire to attain a stipulated temperature, using information given, for example, in References 2 and 3, a knowledge of the factor K will readily give the modified values of current I_2 which will give the same wire temperature at a given position in a filament structure. Thus, for any short length of wire over which the temperature may be considered uniform, the power equations are, for a single wire,

$$I_1^2R = \sigma\phi AT^4 \quad . \quad . \quad . \quad . \quad . \quad . \quad (1a)$$

and for the same wire in a structure,

$$I_2^2R = K\sigma\phi AT^4 \quad . \quad . \quad . \quad . \quad . \quad (1b)$$

Hence $$I_2^2R = KI_1^2R$$

where R is the wire resistance per unit length.

Since R is the same for both conditions at the same temperature,

$$I_2 = \sqrt{K}I_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In any filament structure of finite length, K varies along the wire. It is shown in Sections 3 that although K may vary (normally between values of 0·7 and 1·0), T will vary only to a very small extent. From these considerations it can be shown that in practice, for tungsten or molybdenum filaments at temperatures in the region of 2000° K, conduction effects along the wire are negligible, provided that the length of the strand is large compared with its diameter, as it invariably is in modern transmitting-valve filaments. The problem then, for practical purposes at such temperatures, is one of heat radiation only.

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Pohl is with Standard Telephones and Cables, Ltd.

(3) TEMPERATURE VARIATION FOR TUNGSTEN FILAMENTS

It was shown that the power radiated per unit length is for l practical purposes proportional to KT^4. For constant current ıe power input per unit length is proportional to the resistivity hich, according to Clark and Neuber,[4] is proportional to $(T)^{1\cdot 21}$. ence

$$T^{1\cdot 21} \propto KT^4, \text{ i.e. } KT^{2\cdot 79} \text{ is constant.}$$

If T_b is the temperature at the ends of the structure and T_c the temperature at the centre of the structure, and K_b and K_c ·e the corresponding values of the factor K, then

$$K_bT_b^{2\cdot 79} = K_cT_c^{2\cdot 79} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$T_b = T_c\left(\frac{K_c}{K_b}\right)^{0\cdot 358} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3a)$$

$$T_c = T_b\left(\frac{K_b}{K_c}\right)^{0\cdot 358} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3b)$$

his enables the centre temperature to be found if a filament as been designed for a stipulated temperature at the end of the fective emitting area, or vice versa.

(4) CYLINDRICAL FILAMENT-STRUCTURE OF INFINITE LENGTH

A cylindrical structure of infinite length will first be condered. If the wires are at a uniform temperature the power ow is outwards in a plane perpendicular to the axis of the ructure (see Fig. 2). This power flow can be considered to be dependent of θ (as defined in Fig. 2) if the number of wires is rge, and if the structure is viewed from a distance large comıred to its diameter. For this case, the factor K, which may be ılled K_∞, is given by

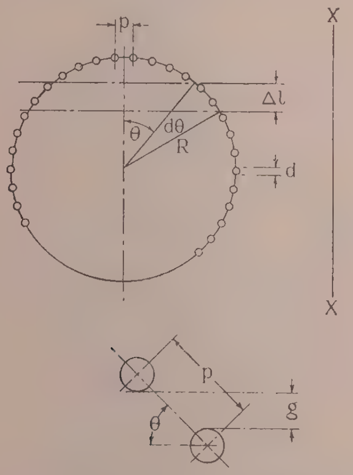


Fig. 2.—Cross-section of filament structure of infinite length.

$$_\infty = \frac{\text{Intensity of radiation at point Z a very large distance from the structure}}{N \times \text{Intensity due to each individual wire at this point}}$$

here N is the number of wires in the structure. The numerator this expression is proportional to the area per unit length of ructure when viewed from Z, while the denominator is proportional to $N\pi d$. These facts are used in Section 10 to show that K_∞ is given by the expression

$$K_\infty = \frac{1}{\pi}\left[\pi - 2\alpha - \frac{d}{p}\log_e\left(\frac{1}{\tan\alpha} + \frac{1}{\sin\alpha}\right) + \frac{p}{d}(1 - \cos\alpha)\right]. \quad (4)$$

in which

d = Wire diameter
p = Circumferential pitch
α = arc sin d/p

This function is plotted in Fig. 3.

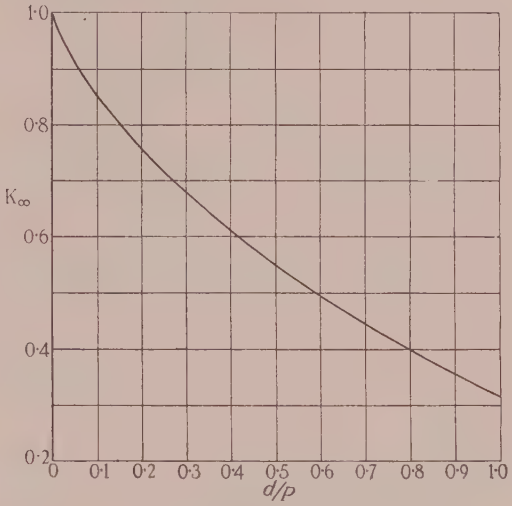


Fig. 3.—Values of K for infinitely long structures.

Using similar methods it is readily shown that for a thin-tape structure the relevant expression is

$$K_\infty = 1 - \frac{w}{2p} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where w is the width of the tape, and p is the pitch.

It should be realized that eqn. (4) has been developed by methods which assume that the number of wires is very large, and for small numbers an appreciable error may result. Using a different approach it is possible to assess this error, but only for values of d/p less than about $0\cdot 5$ can this be done accurately. The procedure is as follows.

Consider an individual wire P, which from symmetry will be representative of any of the wires. Since all the wires can be considered as radiating heat uniformly in all directions perpendicular to their axes, each wire other than P will receive from P an amount of heat per second equal to the amount returned to P by each wire. Therefore for this case the factor K_∞ is given by

$$K_\infty = \frac{1}{2\pi}\left(2\pi - \begin{array}{c}\text{Angle subtended at P}\\ \text{by other wires}\end{array}\right) \quad . \quad . \quad (6)$$

Let R be the radius of the structure. Then if d is the wire diameter and N is the number of wires

$$K_\infty \quad \frac{1}{2\pi}\left\{2\pi - \frac{d}{2R}\left[\frac{1}{\sin\frac{\pi}{N}} + \frac{1}{\sin\frac{2\pi}{N}} + \frac{1}{\sin\frac{3\pi}{N}} + \ldots \frac{1}{\sin\left(\pi - \frac{\pi}{N}\right)}\right]\right\} \quad . \quad . \quad . \quad . \quad (6a)$$

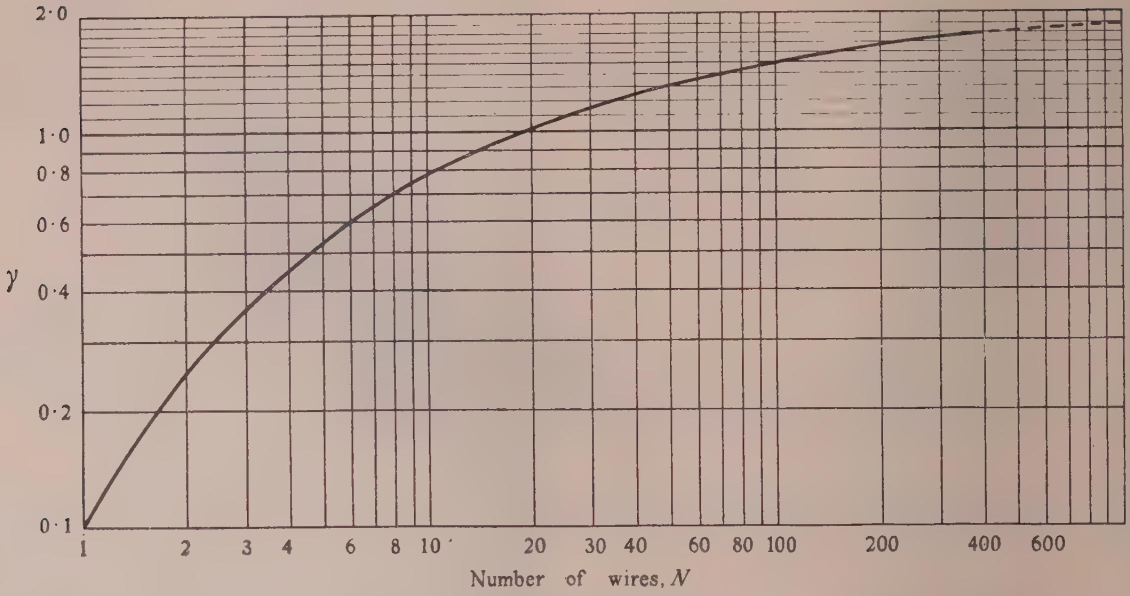


Fig. 4.—Values of function γ used in eqn. (7).

Putting

$$\gamma = \frac{1}{2N}\left[\frac{1}{\sin\frac{\pi}{N}} + \frac{1}{\sin\frac{2\pi}{N}} + \frac{1}{\sin\frac{3\pi}{N}} + \ldots \frac{1}{\sin\left(\pi - \frac{\pi}{N}\right)}\right]$$

and since
$$\frac{N}{R} = \frac{2\pi}{p}, \quad K_\infty = 1 - \frac{d}{p}(\gamma) \quad . \quad . \quad . \quad . \quad (7)$$

Values of γ are plotted in Fig. 4 against N, and it is clearly seen that if N is greater than about 100, γ approaches an asymptotic value, and for such cases it is best to use eqn. (4) because the method used in deriving eqn. (7) makes no allowance for possible overlap of the angles subtended at P by the wires nearest to it. Eqn. (7) may be readily corrected to take this into account for any particular case, but in most practical instances for conditions where eqn. (7) is valid (e.g. where d/p is less than 0·2), the effect is negligible.

When $N = 100$, and $d/p = 0{\cdot}1$ the results from eqns. (7) and (4) agree within 1%. It is suggested, therefore, that eqn. (7) be used where the number of wires is less than 100, and eqn. (4) where it exceeds this figure.

If N is very large, K_∞ may be found by putting

$$K_\infty = \frac{1}{2\pi}\text{(Total angle subtended at P by gaps between wires)}$$

$$= \frac{1}{2\pi}\left(2\int_\alpha^{\pi/2}\frac{p\sin\theta - d}{p\sin\theta}d\theta + \pi - \frac{d}{p}\right)$$

$$= \frac{1}{\pi}\left[\pi - \alpha - \frac{1}{2}\frac{d}{p} - \frac{d}{p}\log_e\left(\frac{1}{\tan\alpha} + \frac{1}{\sin\alpha}\right)\right] \quad . \quad . \quad (7a)$$

For small values of d/p this becomes

$$K_\infty = \frac{1}{\pi}\left[\pi - \frac{d}{p}\left(\frac{3}{2} + \log_e\frac{2p}{d}\right)\right] \quad . \quad . \quad . \quad (7b)$$

which is the same result as that obtained in Section 10 by a different method.

(5) THREE-DIMENSIONAL TREATMENT

It is convenient first to consider a tape structure as shown in Fig. 5. The problem of the infinite structure can be solved in three dimensions by considering a small area, dA, at the point Q on the inside surface of one of the tapes. The amount of heat returned to this small area by the rest of the structure could be summed by the use of well-known fundamental cosine-law relationships (see, e.g., Reference 5, pages 52 and 53). Rather than integrate the effect of the other tapes along their length, a somewhat different method will be used here, in order that the

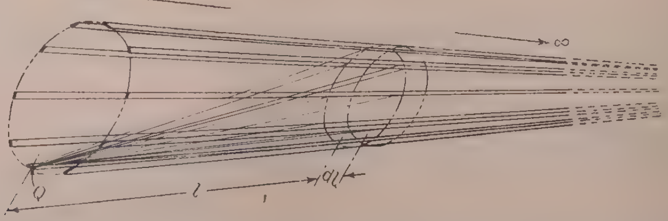


Fig. 5.—Perspective view of tape-filament structure for three-dimensional treatment.

asoning may be readily applicable to circular wire as well as pe structures. Consider a small length dl of the cylinder, at a stance l from Q, as shown in Fig. 5. If ΔP_1 is the power hich the small area at Q radiates into the dark spaces, i.e. the nount of power which is not returned in equal measure, and the area dA radiates a total amount of power PdA into space, being the total emissive power, the factor K_∞ would be given by

$$_\infty = 1 - \frac{\text{Power returned}}{\text{Power radiated}} = \frac{\text{Power not returned}}{\text{Power radiated}}$$

$$= \frac{2\int_{l=\infty}^{l=0} \Delta P_1 + PdA}{2PdA} = \frac{P_1}{2PdA} + \frac{1}{2}$$

here P_1 is the sum of the unreturned power radiated by both des of the elemental area. This expression takes into account e fact that the opposite side of the small area at Q on the tape diates an amount PdA but receives no power in return.

Let $\Delta P_1 = g\Delta P$ where the factor g is the unreturned fraction the total power ΔP_1 radiated by the area dA to the cylindrical ement. If the number of wires or tapes is large, this factor g ill be independent of the axial position of the element and erefore constant.

nce $$K_\infty = \frac{\text{Total power not returned}}{\text{Total power radiated}}$$

the centre of an infinite structure, where $P_1 = PdA$ for one ce, and $gP(dA)$ for the other

$$K_\infty = \frac{PdA + gPdA}{2PdA} = \frac{1+g}{2}$$

ence $$g = 2K_\infty - 1$$

This relation will be used in Section 6, where it is of value ecause it is independent of the length or diameter of the ructure.

(6) STRUCTURES OF FINITE LENGTH

Consider a structure of finite length L for which it is desired calculate the value of K at the centre, i.e. the point Q is at $= L/2$. For both sides of the small area at Q, each side radiating ower equal to PdA, the power which is not returned is

$$P_1 = 2gdA\int_{l=0}^{l=L/2} \Delta P + 2dA\int_{l=L/2}^{l=\infty} \Delta P + PdA$$

Now $$\int_0^{L/2} \Delta P = \int_0^\infty \Delta P - \int_{L/2}^\infty \Delta P$$

that $$P_1 = 2dA\left[g\int_0^\infty \Delta P + (1-g)\int_{L/2}^\infty \Delta P\right] + PdA$$

Now $$\int_0^\infty \Delta P = P/2$$

ence $$K_c = \frac{P_1}{2PdA} = \frac{g}{2} + (1-g)\left(\frac{1}{P}\int_{L/2}^\infty \Delta P\right) + \frac{1}{2}$$

Since $g = 2K_\infty - 1$ this becomes

$$K_c = K_\infty + 2(1 - K_\infty)\left(\frac{1}{P}\int_{L/2}^\infty \Delta P\right) \quad . \quad . \quad . \quad (8)$$

The expression $\frac{1}{P}\int_{L/2}^\infty \Delta P$ is the fraction of the total power radiated by a small area (on the surface of a cylinder distance $L/2$ from the end) which passes through the open end of the cylinder. The calculation is given in Section 10.2, and the results are plotted in Fig. 6, curve (a).

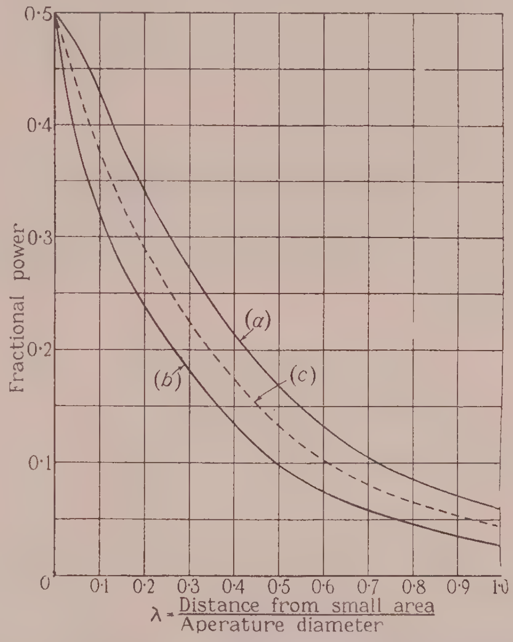


Fig. 6.—Fractional power radiated through one end aperture of an open cylinder from a very small area on the cylinder surface.

For the end of a cylinder,

$$P_1 = dA\left(\tfrac{3}{2}P + g\int_0^L \Delta P + \int_L^\infty \Delta P\right)$$

Proceeding as before, i.e. substituting

$$\int_0^L \Delta P = \int_0^\infty \Delta P - \int_L^\infty \Delta P$$

and $$g = 2K_\infty - 1$$

we obtain $$K_b = \frac{K_\infty}{2} + \frac{1}{2} + (1 - K_\infty)\left(\frac{1}{P}\int_L^\infty \Delta P\right) \quad . \quad . \quad . \quad (9)$$

$\frac{1}{P}\int_L^\infty \Delta P$ is found from Fig. 6.

It is also easy to derive the value of K at any position at a distance x from the end of the cylinder by proceeding in the same manner.

For this it is found that

$$K_x = K_\infty + (1 - 2K_\infty)\left[\left(\frac{1}{P}\int_x^\infty \Delta P\right) + \left(\frac{1}{P}\int_{L-x}^\infty \Delta P\right)\right]$$

Again the expressions $\frac{1}{P}\int_x^\infty \Delta P$ and $\frac{1}{P}\int_{L-x}^\infty \Delta P$ are readily found from Fig. 6.

This reasoning has so far been applied only to a tape structure as shown in Fig. 5. For a circular wire, the surface may be regarded as a large number of elemental surfaces which are either perpendicular or parallel to a tangential plane of the cylinder.* It is therefore necessary also to evaluate the expressions for surfaces perpendicular to the tangential planes, and this is given by curve (*b*) in Fig. 6, the derivation of which is also given in Section 10.2. Now for circular wire, the total area of the infinitesimal surfaces which are perpendicular to the tangential plane will be equal to those which are parallel to it, so that the arithmetic mean between curves (*a*) and (*b*) is applicable. This is given by curve (*c*).

In the foregoing it has been assumed that the filament is at a uniform temperature throughout its length. The error due to this is thought to be small because of the effect of the cosine law of radiation. This causes the heat interchange between the area at Q and a part of the structure at a distance *l* from Q to become rapidly less important in the determination of *K* at Q as *l* increases, i.e. as the temperature begins to differ from that of the area at Q.

(7) SUMMARY OF RESULTS

(7.1) *K* as a Function of Position

For the centre of a cylindrical structure of effective emitting length *L* and diameter *D*, we have from eqn. (8)

$$K_c = K_\infty + 2(1 - K_\infty)f(\lambda) \quad . \quad . \quad . \quad (9a)$$

Here $\lambda = L/2D$ and $f(\lambda)$ is given in Fig. 6, in which curve (*c*) should be used for wire structures, and curve (*a*) for thin tapes.

K_∞ may be found from eqns. (4) or (7). In accordance with the discussion in Section 4, it is suggested that where the number of wires exceeds 100, eqn. (4), i.e. Fig. 3, be used. For less than 100 wires, the use of eqn. (7) and Fig. 4 is recommended, and where thin tape is used instead of wires, eqn. (5) should be used.

For the end of the structure (not the physical end but the end of the effective emitting length)

$$K_b = \frac{K_\infty}{2} + \frac{1}{2} + (1 - K_\infty)f(\lambda) \quad . \quad . \quad . \quad (9b)$$

Here $\lambda = L/D$ and K_∞ is found as before. Again $f(\lambda)$ is given in Fig. 6 for this value of λ.

Although these calculations apply strictly only to filament structures in which the elements are parallel to the axis, they may be used for helical or other cases where the wires or tapes are evenly distributed over the structure. For these, it is suggested that *N*, the number of elements of an "equivalent parallel structure," be taken as

$$N = \frac{\text{Total length of elements in the structure}}{L}$$

$$\text{and } p = \frac{\pi D}{N}, \text{ hence } \frac{p}{d} = \frac{\pi D}{Nd}$$

and values of K_∞ are then obtained from Fig. 3, or eqn. (5) in the case of tape.

That this method is accurate for a filament structure of infinite length is readily shown. For a structure of finite length, the mean effective value of the "equivalent parallel" structure as here defined will also be correct. The substitution in the case of tape structures is easily shown to be equivalent in every way.

(7.2) Example to illustrate Use of the Curves

A cylindrical tungsten-filament structure has an effective emitting length of 2 in, and is 2 in in diameter. The total length of wire in the structure is 420 in, and the wire diameter is 0·006 in.

Given that the temperature at the end of the effective emitting length is to be 2000° K, it is required to find

(*a*) The required current per wire.
(*b*) The temperature in the centre of the structure.
(*c*) The percentage power saved as a result of mutual heating if the temperature at the end of the effective emitting length is to be 2000° C.

$$p \text{ is given by } \frac{\pi D}{N}, \text{ where } N = \frac{\text{Total length of wire}}{\text{Structure length}}$$

$$N = \frac{420}{2} = 210 \text{ in}$$

$$P = \frac{\pi \times 2}{210} = 0{\cdot}030 \text{ in}$$

Hence $\dfrac{d}{p} = \dfrac{0{\cdot}006}{0{\cdot}030} = 0{\cdot}2$; also $\dfrac{D}{L} = 1{\cdot}0$

For the centre of the structure eqn. (9*a*) is used.

$$K_c = K_\infty + 2(1 - K_\infty)f(\lambda)$$

K_∞ is given in Fig. 3, and for $d/p = 0{\cdot}2$ $K_\infty = 0{\cdot}75$.
$\lambda = L/2D = 0{\cdot}5$. $f(\lambda)$ from Fig. 6, curve (*c*) (circular wire) is 0·132.

Hence $K_c = 0{\cdot}75 + 2(0{\cdot}25) \times 0{\cdot}132 = 0{\cdot}75 + 0{\cdot}066 = 0{\cdot}81$.

For the effective end of the structure, K_b is given by eqn. (9*b*),

$$K_b = \frac{K_\infty + 1}{2} + (1 - 2K_\infty)f(\lambda)$$

where for this case $\lambda = L/D = 1{\cdot}0$.

$f(\lambda)$ from Fig. 6 is 0·045.

Hence $K_b = \dfrac{0{\cdot}75}{2} + \dfrac{1}{2} + 0{\cdot}25 \times 0{\cdot}045 = 0{\cdot}887$.

From Reference 2 for a straight wire at 2000° C the current per wire is found to be 1·95 amp for wire of 0·006 in diameter.

Taking into account mutual heating, in order to sustain the same temperature at the end of the effective emitting length the current must be multiplied by $\sqrt{K_b}$.

Hence the current required is

$$1{\cdot}95 \times \sqrt{0{\cdot}887} = 1{\cdot}95 \times 0{\cdot}94 = 1{\cdot}835 \text{ amp}$$

The centre of the filament will run at a temperature given by eqn. (3*b*).

$$T_c = 2000\left(\frac{K_b}{K_c}\right)^{0{\cdot}365} = 2000\left(\frac{0{\cdot}887}{0{\cdot}81}\right)^{0{\cdot}365} = 2070° \text{C}$$

The saving in power is very nearly proportional to K_b. Expressed as a percentage it is

$$\frac{1{\cdot}0}{0{\cdot}887} \times 100 = 11{\cdot}4\%$$

(8) ACKNOWLEDGMENT

The author is indebted to Standard Telephones and Cables, Ltd., for permission to publish the paper, and to Mr. D. C. Rogers and Mr. A. W. Gent for helpful criticisms.

* Since the factor *K* is independent of emissivity, such a substitution is permissible for the purposes of this calculation.

(9) REFERENCES

1) AYER, R. B.: "The Use of Thoriated Tungsten Filaments in High-Power Transmitting Tubes," *Proceedings of the Institute of Radio Engineers*, 1952, **40,** I, p. 591.

2) JONES, H., and LANGMUIR, I.: "The Characteristics of Tungsten Filaments as Functions of Temperature," *General Electric Review*, 1927, **30,** No. 6, p. 310.

3) DAILEY, H. J.: "Designing Tungsten Filaments," *Electronics*, January, 1948, **21,** p. 107.

4) CLARK, J. W., and NEUBER, R. E.; "End Cooling of Power Tube Filaments," *Journal of Applied Physics*, November, 1950, **21,** 1084.

5) MCADAMS, W. H. "Heat Transmission" (McGraw-Hill Book Co., 2nd edition).

(10) APPENDIX

(10.1) Calculation of K_∞ for Structures with Circular Wires

Reference should be made to Fig. 2.

$$K_\infty = \frac{I}{I_2} = \frac{\text{Intensity of radiation at a very large distance}}{N \times \text{Intensity at the same distance due to an indiv. wire}}$$

vhere N is the number of wires, of diameter d. The denominator)f this expression, I_2, is equal to CNd, where C is a constant. The numerator is given by $C \times$ area per unit length of shadow vhich parallel light would throw on a plane such as XX parallel o the axis of the structure, and is therefore

$$I_1 = C\left[2\int_{\alpha}^{\pi n} \frac{nd}{\Delta l}\left(1 + \frac{ng}{\Delta l}\right)dl + 2R(1 - \cos\alpha)\right]$$

Where $n =$ Number of wires in a small arc AB
$\Delta l =$ Projection of AB on XX
$g =$ Projection of the gap between two wires on XX.
$p =$ Circumferential pitch $= \dfrac{2\pi R}{N}$
$R =$ Radius of structure.

$\alpha =$ Angle at which $g = 0$. Since $g = p \sin\theta - d$, $\alpha =$ rc sin d/p. Since $\Delta l = np \sin\theta = R \sin\theta d\theta$, we have

$$_1 = 2CR\left[\frac{2d}{p}\left(\frac{\pi}{2} - \alpha\right) - \frac{d^2}{p^2}\log_\varepsilon\left(\frac{1}{\tan\alpha} + \frac{1}{\sin\alpha}\right) + 1 - \cos\alpha\right]$$

Now $$I_2 = CNd = \frac{2CR\pi d}{p}$$

Hence

$$K_\infty = \frac{1}{\pi}\left[\pi - 2\alpha - \frac{d}{p}\log_\varepsilon\left(\frac{1}{\tan\alpha} + \frac{1}{\sin\alpha}\right) + \frac{p}{d}(1 - \cos\alpha)\right] \quad (4)$$

This expression is plotted in Fig. 3.

Suppose d/p is small, so that $\alpha = \sin\alpha = \tan\alpha = d/p$.

Then $$K_\infty = \frac{1}{\pi}\left[\pi - \frac{d}{p}\left(\frac{3}{2} + \log_\varepsilon\frac{2p}{d}\right)\right]$$

Comparing eqn. (4) with eqn. (7*a*) it may be shown that the lifference is very nearly $1/8\pi(d/p)^3$. Since eqn. (4) may be conidered accurate, this is a measure of the error introduced by he assumptions in deriving eqn. (7). If d/p is, for example, 0·5, he error is $(1/8\pi)(0{\cdot}5)^3 = 0{\cdot}005$, so that for practical purposes qn. (7) is accurate within 1% if d/p is less than 0·5.

(10.2) Fraction of Total Power radiated through the Open End of a Cylinder of Diameter *D* by a Small Area *dA* on the Curved Surface, at a Distance *L* from the open end. (See Reference 5, pp. 52 and 53.)

(10.2.1) Small Element *dA*, Parallel to the Cylindrical Surface.

Referring to Fig. 7(*a*), the fraction *dP* of the total power radiated by *dA*, which is received by a small area dA_2, of height *dy* and width *dx*, in a plane perpendicular to dA_1 is readily shown to be

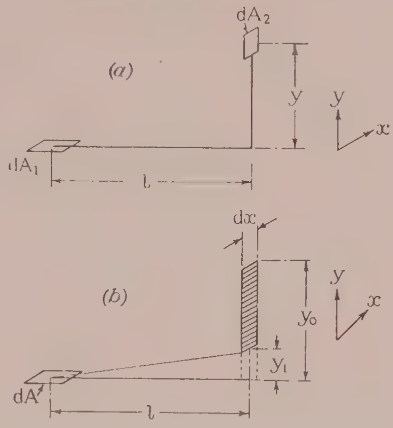


Fig. 7.—Co-ordinates and symbols used in Section 10.2.

$$dP = \frac{dxdy}{\pi}\frac{ly}{(l^2 + y^2)^2}$$

Hence for an elemental strip width, extending from y_1 to y_0 [Fig. 7(*b*)]

$$\int dP = \Delta P = \frac{ldx}{\pi}\int_{y_0}^{y_1}\frac{ydy}{(l^2 + y^2)^2} = \frac{ldx}{2\pi}\left[\frac{y_1^2 - y_0^2}{(l^2 + y_1^2)(l^2 + y_0^2)}\right]$$

The shaded strips in Fig. 8(*a*) must be integrated to cover the circular end of the cylinder shown, so that

$$y_1 = 1 + \sqrt{(1 - x^2)} \quad y_0 = 1 - \sqrt{(1 - x^2)} \quad \cos\psi = L/l$$

Here all dimensions are normalized with respect to the radius of the cylinder.

$$P = 2\int_{x=0}^{x=1}\Delta P \cos\psi$$

where ΔP is given above.

Substituting for ΔP and inserting the values for y_1, y_0, $\cos\psi$, putting $l^2 = L^2 + x^2$, and simplifying,

$$P = \frac{L}{\pi} \times$$

$$\int_0^1 \frac{4\sqrt{(1 - x^2)}dx}{\{L^2 + x^2 + [1 - \sqrt{(1 - x^2)}]^2\}\{L^2 + x^2 + [1 + \sqrt{(1 - x^2)}]^2\}}$$

This integral has been evaluated and the results are shown in Fig. 6 curve (*a*), where *P* is plotted as a function of the normalized length

$$\lambda = \frac{\text{length}}{\text{diameter}} = \frac{L}{2}$$

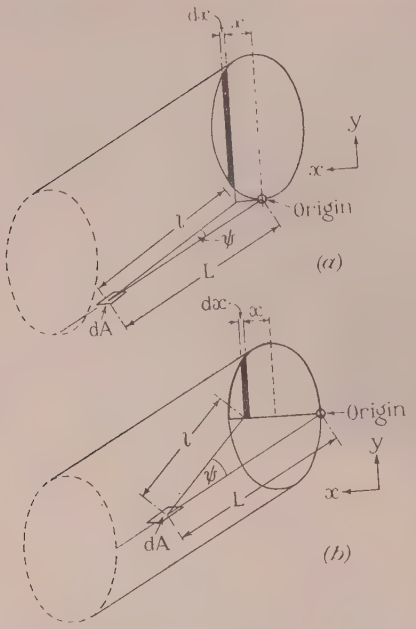


Fig. 8.—Co-ordinates and symbols used in Section 10.2.

That $P = 0{\cdot}5$ when $\lambda = 0$ is consistent with conclusions drawn from purely physical considerations, as is the fact that the curve is asymptotic to the abscissa.

(10.2.2) Small Element dA, in a Radial Plane, i.e. Perpendicular to the Surface.

The plane of the element dA is perpendicular to that of the previous case [see Fig. 8(*b*)].

Procedure is the same as in Section 10.2.1. In this case all dimensions are normalized with respect to the diameter of the cylinder, and thus we obtain

$$y_1 = \sqrt{(x - x^2)} \qquad y_0 = 0$$

In order that eqns. (8) and (9) may still be correct, we need to insert a factor of 2 because here both sides of dA radiate through one half of the open end.

$$P = 2\int_0^1 \Delta P \cos \psi$$

Using again the expression for ΔP and substituting $y_1 = \sqrt{(x - x^2)}$, $y_0 = 0$, $\cos \psi = L/l$ we obtain

$$P = \frac{L}{\pi}\int_0^1 \frac{(x - x^2)dx}{(L^2 + x^2)(L^2 + x)} = \frac{L}{\pi}\int_0^1 \left(\frac{1}{L^2 + x^2} - \frac{1}{L^2 + x}\right)dx$$

$$P = \frac{\text{arc tan } 1/L}{\pi} - \frac{L}{\pi}\log_\varepsilon\left(1 + \frac{1}{L^2}\right)$$

Again, when $L = 0$ $P = 0{\cdot}5$, which is what would be expected.

This function is plotted in Fig. 6, curve (*b*). Fig. 6, curve (*c*), shows the mean value of curves (*a*) and (*b*), which is applicable to wire elements of square or circular cross-section.

In Fig. 6, λ equals L for these conditions, since L had been normalized with respect to the diameter.

DISCUSSION ON
"A NOTE ON THE SURFACE LOSS IN A LAMINATED POLE-FACE"*

Mr. K. C. Mukherji (*communicated*): In the course of an attempt to obtain a general formula applicable to tooth-ripple losses and losses due to load harmonics in electrical machines, I have been engaged for some time in an investigation of cylindrical electromagnetic waves of various peripheral wavelengths, rotating in space with various angular velocities, such as those existing in the air-gap of most rotating machines. The author's paper is therefore of great interest to me. My approach to the problem, although independently developed some time before the author's work was published, had in fact been essentially similar to his, with limitations similar to those implicit in his paper, namely that all discontinuities in the volume or surface of the pole due to slots or wedges were neglected, and that the air-gap field was assumed to remain unaffected by the field produced by the eddy currents. However, I obtained my solutions in terms of the cylindrical polar system of space co-ordinates, involving modified Bessel functions of complex arguments. Fortunately, the arguments were large enough in the case of the tooth-ripple harmonics to justify the use of the asymptotic expansions of these functions at the final stage, and in so doing, my result degenerated exactly into the author's eqn. (20).

Since then, however, a closer review of the work, prompted by some experimental observations, has revealed two important points relating to my solutions of Maxwell's equations, or equally to those obtained by the author. These are, first, that the solutions are not wholly consistent physically, and secondly, that they are not sufficient to explain the entire phenomenon of tooth-ripple flux pulsation as observed in a particular laminated pole-shoe. I should like to elaborate these two points as follows:

(*a*) It will be noted from the author's eqn. (8) that B_x is an odd function of x and has finite values at $x = \pm h/2$; this implies a discontinuity in the axial component of the induction at the boundaries between adjacent laminations in a laminated pole-shoe. There is nothing to be said against such an expression for B_x when only a single lamination is considered by itself, but it is surely inadmissible in the case of the usual compact laminated structure.

(*b*) A group of experiments on the penetration of tooth-ripple flux pulsation into a laminated pole-face has been made on an experimental machine with stationary salient poles employing

* CARTER, G. W.: Monograph No. 123, March, 1955 (see 102 C, p. 217).

·18 mm cold-rolled close annealed pole-stampings, and an unvound rotor having a tooth-pitch of 24·4 mm. Axial holes of /16 in diameter were drilled in the pole stampings, arranged in airs one-half tooth-pitch apart circumferentially, each pair being paced radially at successive depths of 1·5 mm, 3·0 mm, and ·0 mm behind the pole-face. The holes in individual stampngs were aligned so as to allow the threading of a search-coil vire throughout the full length of the pole, two such wires orming a single-turn coil one-half tooth-pitch wide. An dditional search coil of the same width was glued to the ole-face. A third search coil was wound embracing the entire ole-face with a view to observing the main flux pulsation. This vas found to be relatively small. Measurements of the e.m.f.'s nduced in the other search coils, using an electronic wavenalyser, showed that the attenuation of the flux-ripples took lace, in fact, at a much lower rate than that envisaged by the uthor's eqn. (23), which should have been valid for our paricular experiment; still more striking was the observation that his rate of attenuation was independent of the speed of the otor. In fact, the measurements suggested an exponential lecay of the flux-ripples, roughly according to the law $\varepsilon^{-2\pi z/q}$.

Rough calculation indicated that holes of the size used would ot affect the induced e.m.f.'s by more than about 15%. We herefore had to look back to Maxwell's equations again for an xplanation of this apparent anomaly, and it became clear that a olution of the equations of the form

$$\cosh\left(\sqrt{(j)}\frac{x}{d}\right)\varepsilon^{-\frac{2\pi}{q}z-j\frac{2\pi}{q}(y-vt)}$$

vas not only perfectly feasible but also possibly compatible with nore exact boundary conditions of the problem. A rather xtensive mathematical investigation of the entire phenomenon s now in progress in association with Prof. H. Bondi.

The field variation across the thickness of the laminations, as nvisaged by this particular type of solution, is familiar to us in the classical problem of the field distribution in an infinite lamina of restricted thickness placed in the alternating field of an infinitely long solenoid. In fact, the existence of this phenomenon in association with the tooth-ripple flux pulsation in laminated poles was referred to (if only qualitatively) by Adams[*] at the beginning of this century, when he described the circumferential elements of pole-face eddy currents as "screening currents" because of their tendency to screen the centre of the pole-face from the tooth-ripple flux pulsations.

Prof. G. W. Carter (*in reply*): I was very interested to learn that Mr. Mukherji and I had been working on similar lines, and that he had obtained a formula for the surface loss perhaps earlier than I had. I did not consider it worth while, however, to take account of the cylindrical curvature of the surfaces, in view of the many other, more serious differences between the idealized problem and an actual motor.

The finite value of the normal flux-density at the faces of a lamination, explicit in eqn. (8) of my paper, is implicit also in my father's work. Such a component of flux-density would arise, in a single lamination, from the magnetic effect of the eddy currents themselves. It is therefore probable that the turning of a blind eye to this component (which would not exist if the lamination were in the middle of a pile) is of a piece with the general neglect of the field of the eddy currents in comparison with the inducing field. This neglect is admittedly an imperfection in my solution; it is therefore gratifying to learn that Mr. Mukherji believes himself to have discovered another solution which promises to be less open to criticism. Further particulars will be welcome, for it is not easy to see how the frequency of the disturbance and the physical properties of the material, summarized in the constant d, can be without effect on the attenuation in the z-direction, yet can make their appearance in the mode of variation of the quantities in the x-direction even on the plane $z = 0$.

[*] *Transactions of the American I.E.E.*, 1909, **28**, p. 1133.

DISCUSSION ON
"STEADY-STATE STABILITY OF SYNCHRONOUS GENERATORS AS AFFECTED BY REGULATORS AND GOVERNORS"[*]

Mr. D. Broadbent (*Australia: communicated*): The paper is a progressive step in the modern treatment of power systems and their regulators as closed-loop systems with their components representable by transfer functions. So long as the operation is restricted so that the systems can be represented by linear equations, the method has much to recommend it. However, the parameters of machines and loads are not constant, as the authors have indicated, and an exact mathematical treatment would be very complicated, particularly for paralleled machines and loads. For specific problems a miniature machine analogue[†] is simpler.

These remarks apply equally to the problems in governor operation. Various authors, including Crary, have suggested that the low speed of operation of the governor loop justifies taking the prime-mover power as a constant for the first swing of the stability study. An approach of the type indicated in the paper is the means for confirming or modifying this. Unfortunately, while the authors have gone to some trouble to introduce the effects of machine field time-constant by τ'_{d_z} in eqn. (16), they have neglected the all-important τ_1 and τ_2 in their solutions. This omission has the effect of increasing the damping coefficient from 4·0 to about 24·0 without raising the order of the equation. In fact, because of its inherent time-constants, and its dead-band, a speed-error governor gives nothing like this degree of damping and may even excite oscillations.

Partly for this reason a time-error governor[†] with stabilizing circuits[‡] was developed at Melbourne University Electrical

[*] Messerle, H. K., and Bruck, R. W.: Monograph No. 134 S (see page 24).

[†] Mackley, K. W.: "Development of Model Power System," *Electrical Engineer (Melbourne)*, 1955, **32**, p. 117.

[†] Broadbent, D.: "Integral Governing of Turbo-Alternators," *Electrical Engineer (Melbourne)*, 1953, **29**, p. 354.

[‡] Broad bent, D.: "Stability of Integral Governing," *ibid.*, 1955, **32**, p. 40.

Engineering Department for use in a miniature machine system. Tests were made using a differential analyser* for an isolated machine and two paralleled machines time-error governed.

Eqn. (A) describes the isolated machine behaviour, the nomenclature used being the same as that in the paper.

$$Mp^2\Delta\delta + Dp\Delta\delta + \Delta T_m = 0 \quad . \quad . \quad . \quad \text{(A)}$$

where $\Delta T_m = \dfrac{K\Delta\delta}{(\tau_1 p + 1)(\tau_2 p + 1)}$ for simple time-error governing.

K = Governor gain
= 0·02 per unit in the tests.

The difference in the solutions using $\tau_1 + \tau_2 = 0$ and a design figure of $\tau_1 + \tau_2 = 0{\cdot}3$ sec was negligible. This is partly because their omission in a time-error-governed system does not introduce heavy damping.

Eqns. (B) and (C) describe two machines connected by a tie-line having a damping coefficient D_{12} and a synchronizing torque coefficient T_{s12}.

$$M_1p^2\Delta\delta_1 + D_1p\Delta\delta_1 + D_{12}p(\Delta\delta_1 - \Delta\delta_2) + T_{s12}(\Delta\delta_1 - \Delta\delta_2) + \Delta T_{m1} = 0 \quad . \quad \text{(B)}$$

$$M_2p^2\Delta\delta_2 + D_2p\Delta\delta_2 + D_{12}p(\Delta\delta_2 - \Delta\delta_1) + T_{s12}(\Delta\delta_2 - \Delta\delta_1) + \Delta T_{m2} = 0 \quad . \quad \text{(C)}$$

$$\Delta T_{m1} = \frac{K_1\Delta\delta_1}{(\tau_{11}p + 1)(\tau_{21}p + 1)}$$

and

$$\Delta T_{m2} = \frac{K_2\Delta\delta_2}{(\tau_{12}p + 1)(\tau_{22}p + 1)}$$

for simple time-error governing.

For tests described $\tau_{11}, \tau_{21}, \tau_{12}$ and τ_{22} were neglected, resulting in the governor gain-constant K increasing the effective value of the synchronizing torque coefficient, possibly raising the stability limit. For this measure to be effective, not only must $\tau_1 + \tau_2$ be small, but the value of K must be comparable to T_s. This raises the question of governor stability for the isolated machine. For this reason in the machine analogue the governor incorporated a second-derivative stabilizing feedback according to eqn. (D).

$$\Delta T_m = \frac{1}{(\tau_1 p + 1)(\tau_2 p + 1)}\left(K\Delta\delta + \frac{K^1p^2\Delta\delta}{\tau' p + 1}\right) \quad . \quad \text{(D)}$$

Time-error-governed machines using a low value of K have worked in parallel in a power system and have been described.

Messrs. H. K. Messerle and **R. W. Bruck** (*in reply*): Mr. Broadbent's comments amplify the importance of the effects of governors on the performance of synchronous machines. The governor time-delay modifies the transient response and stability limit, and a more detailed analysis can be found in a later publication.*

Normal integrated speed controllers or time-error governors have a negligible effect so far as fast machine transients are concerned, and the steady-state and dynamic stability limits are only slightly modified. The sensitivity, K, of this type of controller, as implied by Mr. Broadbent, could be increased beyond the values which are normally used. This, however, is usually considered as impracticable for two reasons: first, fast integral control makes the operation of the alternator very unstable, and secondly, there is no point in trying to force large machines or power stations to correct for every load change. In general, the sensitivity is chosen so that the controller averages out the overall speed variations over a period of, say, 30 sec or more and then acts accordingly. By that means excessive machine oscillations are avoided.

The simplified approach for the differential analyser study by Broadbent has been used also by Concordia and Kirchmayer (Reference 13). It neglects the effect of alternator field time-constant, which is very critical in stability studies.*

* BROADBENT, D.: "The Stability of Time-Error Governed Turbines in Power Systems," *Australian Journal of Applied Science*, 1955, 6, p. 281.

* MESSERLE, H. K.: "Relative Dynamic Stability of Large Synchronous Generators," *Proceedings I.E.E.*, Monograph No. 159 S, January, 1956 (103 C),

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

PART C—MONOGRAPHS, MARCH 1956

CONTENTS

Declaration on Fair Copying.—Within the terms of the Royal Society's Declaration on Fair Copying, to which The Institution subscribes, material may be copied from issues of the *Proceedings* (prior to 1949, the *Journal*) *which are out of print and from which reprints are not available.* The terms of the Declaration and particulars of a Photoprint Service afforded by the Science Museum Library, London, are published in the *Journal* from time to time.

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part. *Example.*—SMITH, J.: "Reflections from the Ionosphere," *Proceedings I.E.E.*, Paper No. 3001 R, December, 1954 (102 B, p. 1234).

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* Temple Bar 7676. *Telegrams:* "Voltampere, Phone, London."
Printed by Unwin Brothers Limited, Woking and London.